JERREN GOULD. Automata in Environments (Under the direction of EDWARD J. WEGMAN).

Automata theory is the study of mathematical models of finite-state systems which admit certain inputs and emit certain outputs. The state transition behavior of previously considered models, such as Turing machines, finite automata, and probabilistic automata, has been assumed to be related only to the present input. In this dissertation we take the viewpoint that every system is situated within an environment which affects the initial state distribution and the state transition function. The Rabin-Paz model of a probabilistic automaton is generalized so that the initial state distribution is related to the initial configuration of the environment and that the state transition function is not only related to the present input but also to the present configuration of the environment.

Chapter 2 considers the case in which there is a deterministic rule to specify the configuration of the environment at any time. We show that such a model is properly more general and more powerful than the probabilistic automaton model. Sufficient conditions are given for an automaton with a deterministic environment rule to be "simulated" by a probabilistic automaton. We also consider the behavior of automata with environment sets that have certain metric or semi-group structure.

In Chapter 3 we consider the effects of small perturbations of the environments on the system. Conditions are given for automata in environments to have stable state distribution functions. We also give conditions for the set of tapes defined by an automaton to remain unchanged for sufficiently small perturbations of the environments.
Finally, we consider the case of automata operating in random environments; we assume that the realization of any environment is governed by some probabilistic structure. The behavior of systems with various environment processes is investigated. We encounter the dual nature of the randomness involved in automata in random environments. We have that the environments are random and that for each value the environment assumes, the probabilistic state transitions are defined. However, we find that there is a mean equivalent canonical representation for automata in random environments which eliminates the randomness due to probabilistic state transitions; for any automaton in random environments we can find another automaton in random environments with finite environment set, deterministic assignment of the initial state, and deterministic transitions which has a state distribution equivalent in the mean.
ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Professor Edward J. Wegman for proposing this problem and for the illuminating suggestions he made during the course of the investigation. I am greatly indebted to him for his guidance and encouragement throughout my graduate career. It has been a very valuable experience to work under his direction.

I would also like to thank Professor W.L. Smith, the chairman of my dissertation committee, for his helpful suggestions and for his encouragement during my graduate study. Thanks are also due to Professors N.L. Johnson, G.D. Simons, S. Cambanis, C.A. Magó, and W.R. Wogen for participating as members of my dissertation committee.

The financial support provided by the Department of Statistics during my period of study in Chapel Hill is greatly appreciated.

The contributions of June Maxwell, who ably typed the manuscript, are earnestly appreciated.

Finally, I wish to express my heartfelt gratitude to my wife, Marsha, for being My Small Joy.
TABLE OF CONTENTS

Chapter I: Development of Automata Theory

1.1 The Automaton Concept ------------------------------- 1
1.2 Finite Automata-------------------------------------- 3
1.3 Probabilistic Automata------------------------------- 8

Chapter II: Environments

2.1 The Concept of Environments for Automata----------17
2.2 Automata in Deterministic Environments-------------18
2.3 Mathematical Generality of the ADE Model----------20
2.4 Finite Environment Set------------------------------21
2.5 Reducing the Environment Set and the Induced Metric--29
2.6 The Environment Set as a Semi-group---------------32

Chapter III: Stability Results

3.1 The Stability Problem-------------------------------37
3.2 Stability with Tapes of Bounded Length-------------41
3.3 Ergodic Conditions for s-Stability------------------47
3.4 Response Intervals and a-Stability-----------------64
3.5 Round Off Considerations---------------------------69

Chapter IV: Automata in Random Environments

4.1 Basic Notions---------------------------------------71
4.2 Random Environment Sequences of Independent
  Random Variables--------------------------------------75
4.3 Limit Results for Sequences of Environment Processes---77
4.4 A Canonical Representation for Automata in Random
  Environments------------------------------------------81
Chapter V: Implications

5.1 Limit Results for Markov Systems-----------------------------91
5.2 Applications-------------------------------------------------95
5.3 Conclusions and Summary-------------------------------------97

Bibliography------------------------------------------------------100

Index-------------------------------------------------------------102

Notation----------------------------------------------------------104
CHAPTER I

Development of Automata Theory

1.1 The Automaton Concept.

Automata theory is the study of the characteristics of automatic devices. For some time there has been research into the nature of algorithms and the functions computable by following an algorithm. Researchers in automata theory have made an effort to apply and extend these results to more realistic models of physical systems. The automata models have been motivated particularly by the behavior of digital computers, but also by organic nerve networks, information transmission systems, synthesis of man's languages, learning behavior, games between automata, etc. These models are constructed to investigate their mathematical properties as well as to discover the characteristics and capabilities of similar phenomena in the real world.

In considering the theoretical limitation of the type of computation that could be carried out by an automatic calculating device, A.M. Turing [18] constructed an abstract machine capable of following a finite, consistent list of instructions. The Turing machine is a finite-state machine with an associated external memory. The memory has the form of a sequence of squares marked off on a linear tape. The tape is assumed to be infinite in both directions. A Turing machine has a finite alphabet of symbols and a finite internal state set. The machine is capable of five operations. It can read a square of the
tape, erase a square of the tape and write a new symbol on that square, change internal state, move to an adjacent square, and halt. The instructions to the machine are contained in a finite list of quintuples of the form: (present state, symbol read, new state, symbol written, direction of motion). The third, fourth, and fifth components of the quintuple are determined uniquely as functions of the first and second components. Thus, for any symbol read and internal machine state, the next act of the machine is unambiguously specified. Although the tape is regarded as infinite in both directions, when the machine is started, the tape must be blank except for a finite number of squares. The Turing machine is started by placing the machine on a square of the tape and designating an initial state. The output is considered to be the expression on the tape when the machine halts.

Indeed, any computation that can be performed by a deterministic automatic calculating device can also be performed by a Turing machine. This fact led to "Church's thesis," the assumption that every algorithm has a Turing machine which can execute it. Thus, consideration of Turing machines gave rise to results concerning the theory of computable functions and other equivalent logic systems based on algorithms. The Turing machine model in many ways patterns the operation of a digital computer. The finite state set and finite set of quintuples form a "program." Both systems are capable of accepting any finite amount of input. The major drawback to the Turing machine model for the digital computer is its unbounded memory capacity via its infinite tape. This enables the Turing machine model to consider the entire class of computable functions.
In order to simulate more closely the characteristics of digital computers, the Turing machine model was amended by restricting the external memory tape to be of bounded length. This model, called a linear bounded automaton, operates in the same manner as a Turing machine except that the tape contains only a finite number of squares equal to the length of the input. The machine is not allowed to move beyond the bounds of the tape. The new model is not as general as the Turing machine. Linear bounded automata, however, are capable of computing recursive functions that can be handled by a finite, bounded memory, as in the case of digital computers.

1.2 Finite Automata

So far the discussion has centered on the transducer model of machines. That is, a machine acts as a transducer if its main purpose is to generate an output sequence that is a function of the input and the initial state of the machine. However, for the purposes of this paper we shall henceforth consider machines that act as acceptors. Instead of producing an output sequence, acceptors simply answer "yes" or "no" to questions asked of them. The transducer model and the acceptor model are equivalent in their computing capabilities.

We now formulate an acceptor model with a bounded memory capacity. This class of machines is called finite automata. For finite automata the questions take the form of finite sequences from a finite alphabet; such a sequence of symbols is called a tape. The machine itself shall be considered as a black box, capable of reading one symbol at a time from the tape and then advancing to the next symbol on the tape. In order to maintain the generality of the structure of the model,
the concept of a finite internal state set is employed. As with Turing machines and linear bounded automata, this machine is capable of changing states based on the present state and latest input symbol. With finite automata the memory is characterized by the finite internal state set and, hence, is finite and bounded. To obtain the output, or answer, from the machine its state set is partitioned into two classes, the states for which the output is "yes" and the states for which the output is "no." The output is observed when the machine has read all of the symbols on the tape.

Let us now define these notions of finite automata in a more precise mathematical form. Let \( \Sigma \) be the finite input set, the alphabet, and let \( \Sigma^* \), the set of tapes, be the class of all finite sequences of symbols from \( \Sigma \). Let us also include \( \Lambda \), the empty tape, in \( \Sigma^* \). Note that \( \Sigma^* \) together with the operation of concatenation forms a free semigroup generated by \( \Sigma \). We are now prepared to formalize the definition of finite automata.

**Definition 1.2.1.** A finite automaton (FA) over the alphabet \( \Sigma \) is a system \( A = (S, M, s_0, F) \), where \( S \) is a finite nonempty set (the internal states of \( A \)), \( M \) is a mapping \( M : S \times \Sigma \to S \) (the transition function of \( A \)), \( s_0 \in S \) (the initial state of \( A \)), and \( F \subseteq S \) (the set of final, or "yes," states of \( A \)).

For a finite automaton \( A \) the domain of definition of the transition function can be extended from \( S \times \Sigma \) to \( S \times \Sigma^* \) by the recursion as follows:

\[
M(s, \Lambda) = s \quad \forall s \in S
\]

\[
M(s, x\sigma) = M(M(s, x), \sigma) \quad \forall s \in S, \forall x \in \Sigma^*, \forall \sigma \in \Sigma
\]
The interpretation of \( M(s,x) \) is elementary: it is that state of the machine obtained by beginning in state \( s \) and reading through tape \( x \) symbol by symbol, changing states according to the transition function.

**Definition 1.2.2.** The collection of all tapes \( x \in \Sigma^* \) such that \( M(s_0,x) \in F \) is called \( T(A) \), the set of tapes accepted, or defined, by the automaton \( A = (S, \Gamma, s_0, F) \).

**Definition 1.2.3.** The class of all definable sets of tapes, \( T \), is the collection of all sets of the form \( T(A) \), for some finite automaton \( A \). If \( U \in T \), then \( U \) is called a regular event.

In order to characterize the definable sets of tapes, we must make the following definitions.

**Definition 1.2.4.** An equivalence relation \( E \) over the set \( \Sigma^* \) of tapes is right invariant if whenever \( x \in E \), then \( xz \in E \), \( \forall z \in \Sigma^* \).

**Definition 1.2.5.** An equivalence relation \( E \) over the set of tapes is of finite index if there are only finitely many equivalence classes under the relation.

Hence, we now state a fundamental theorem due to A. Nerode.

**Theorem 1.2.1.** Let \( U \in \Sigma^* \). The following are equivalent:

1. \( U \in T \).
2. \( U \) is the union of some equivalence classes of a right invariant equivalence relation of finite index.
3. The explicit right invariant equivalence relation \( E \) defined by
the condition that \( \forall x, y \in \Sigma^* \), \( x \equiv y \) iff \( \forall z \in \Sigma^* \) whenever \( xz \in \mathcal{U} \),
then \( yz \in \mathcal{U} \), and conversely, is an equivalence relation of finite index.

If \( \mathcal{U} \in \mathcal{T} \), then the number of equivalence classes under the relation \( E \) is the least number of internal states of any finite automaton defining \( \mathcal{U} \). Theorem 1.2.1 also leads to some simple facts about the class \( \mathcal{T} \).

**Theorem 1.2.2.** The class \( \mathcal{T} \) is a Boolean algebra.

**Theorem 1.2.3.** The class \( \mathcal{T} \) contains all finite sets of tapes.

We are also interested in characterizing the set \( \mathcal{T}(A) \) when we are not told how the black box \( A \) operates. All that is allowed is to supply input, or "question," the black box and observe the "answers." Rabin and Scott [17] exhibit an effective procedure\(^1\) based only on a knowledge of a bound on the number of states of \( A \) whereby in a finite number of steps it can be decided whether \( \mathcal{T}(A) \) is empty, finite, or infinite. It is also possible to determine an upper bound on the number of tapes in \( \mathcal{T}(A) \) when \( \mathcal{T}(A) \) is finite. Rabin and Scott further discuss this type of result in comparing finite automata.

**Definition 1.2.6.** Two finite automata \( A \) and \( B \) are equivalent iff \( \mathcal{T}(A) = \mathcal{T}(B) \).

**Definition 1.2.7.** The length of any tape \( x \in \Sigma^* \) is the number of symbols in the sequence \( x \). We denote the length of \( x \in \Sigma^* \) by \( L(x) \).

---

\(^1\) An effective procedure is a set of rules which dictate, from moment to moment, precisely how to behave. Effective procedure is a synonym for algorithm.
Theorem 1.2.4. Let $A_1$ and $A_2$ be FA with $n_1$ and $n_2$ internal states, respectively. $A_1$ and $A_2$ are not equivalent iff there exists $x \in \Sigma^*$ such that $L(x) < n_1n_2$ and $x$ is accepted by one machine and not the other.

Rabin and Scott also introduced the model of a nondeterministic finite automaton. In contrast to a finite automaton which has only one way to change state in any particular situation, the nondeterministic machine allows a choice from a set of states.

Definition 1.2.8. A nondeterministic finite automaton over the alphabet $\Sigma$ is a system $A = (S,M,S_0,F)$, where $S$ is a finite non-empty set, $M$ is a function from $S \times \Sigma$ taking values in the set of all subsets of $S$, $S_0 \subseteq S$, and $F \subseteq S$.

Definition 1.2.9. Let $A$ be a nondeterministic finite automaton. The set of tapes $T(A)$ accepted, or defined, by $A$ is the collection of all tapes $x = \sigma_0 \sigma_1 \ldots \sigma_{k-1}$ for which there exists a sequence $s_0, s_1, \ldots, s_k$ of internal states of $A$ such that

1. $s_0 \in S_0$, $s_i \in M(s_{i-1}, \sigma_{i-1})$ for $i = 1, 2, \ldots, k$,
2. $s_k \in F$.

This appears to be a generalization of finite automata; however, Rabin and Scott demonstrated an effective procedure to construct a finite automaton defining exactly the same set of tapes as any given nondeterministic finite automaton.

Definition 1.2.10. Let $A = (S,M,S_0,F)$ be a nondeterministic finite automaton. $D(A)$ is the system $D(A) = (P,Q,P_0,R)$, where $P$
is the set of all subsets of $S$, $Q$ is a function $Q: P \times E \rightarrow P$ such that $Q(p, \sigma) = \bigcup_{s \in p} M(s, \sigma)$, $p_0 = S_0$, and $R$ is the set of all subsets of $S$ containing at least one element of $F$. Clearly, $\mathcal{D}(A)$ is a finite automaton and is equivalent to $A$.

**Theorem 1.2.5.** If $A$ is a nondeterministic finite automaton, then $T(A) = T(\mathcal{D}(A)) \in T$.

1.3 Probabilistic Automata

So far we have assumed that our black box machine operates in an error-free deterministic manner or in a manner equivalent to such a machine. In physical machines, however, malfunctions may occur as a result of mechanical or electrical breakdown. Rabin [16] proposed the model of probabilistic automata to simulate the behavior of machines and systems with certain types of random errors and to investigate the more general class of machines. Here we combine the mathematical definitions of Rabin and Paz [13] to take on the most general form.

**Definition 1.3.1.** Let $M_n$ denote the set of all $n \times n$ stochastic matrices and $V_n$ denote the set of all $n$-dimensional stochastic vectors.

**Definition 1.3.2.** A probabilistic automaton (PA) over the alphabet $\Sigma$ is a system $A = (S, M, \pi_0, F)$, where $S = \{s_1, \ldots, s_n\}$ is a finite set (the set of internal states of $A$), $M$ is a function $M: \Sigma \rightarrow M_n$ (the matrix transition function) such that $m_{ij}(\sigma)$ is the probability of changing to state $s_j$ under input $\sigma$ given that the machine is in state $s_i$, $\pi_0 \in V_n$ (the initial state distribution) and $F \subseteq S$ (the set of final states).

Probabilistic automata are models for systems capable of a finite number of internal states and able to receive inputs $\sigma \in \Sigma$. When in
state \( s_i \) and if the input is \( \sigma \), then the system may go into any one of the states \( s_j \in S \), and the probability of going into state \( s_j \) is \( m_{ij}(\sigma) \). These transition probabilities are assumed to remain fixed and be independent of time and other inputs. Thus, the chance of an "error" on any transition depends only on the present state of the machine and the input symbol read. We now extend the definition of \( M \) to the set of all tapes of finite length, \( \Sigma^* \), by the rule:

if \( x = \sigma_1 \sigma_2 \ldots \sigma_k \), then \( M(x) = M(\sigma_1)M(\sigma_2)\ldots M(\sigma_k) \).

**Definition 1.3.3.** Let \( D(M_n) \) denote the subset of \( M_n \) such that if \( \forall e \in D(M_n) \), then \( m_{ij} = 0 \) or \( 1 \). Also let \( D(V_n) \) be the subset of \( V_n \) such that if \( \forall e \in D(V_n) \), then \( v_i = 0 \) or \( 1 \).

By restricting the transition function \( M \) to take values in \( D(M_n) \) and the initial distribution function \( \pi_0 \) to take values in \( D(V_n) \) we can now see that the class of probabilistic automata includes the class of finite automata.

**Definition 1.3.4.** Let \( A = (S, M, \pi_0, F) \) be a PA. We define the state distribution of \( A \) after input tape \( x \in \Sigma^* \) as \( \pi(x) = \pi_0 M(x) \).

Let \( \eta^F \) be an \( n \)-dimensional column vector such that \( \eta^F_i = \begin{cases} 1 & s_i \in F \\ 0 & s_i \notin F \end{cases} \).

We define the probability that tape \( x \) is accepted by \( A \) as

\[ p(x) = \pi(x) \eta^F = \pi_0 M(x) \eta^F. \]

As with finite automata, probabilistic automata may be used to define sets of tapes. Because of the more general nature of probabilistic automata, these sets of tapes will depend not only on \( A \), but also on an additional parameter, \( \lambda \), called the cut-point.
Definition 1.3.5. Let $A$ be a PA and $\lambda$ be a real number, $\lambda \in [0,1)$. The set of tapes $T(A,\lambda)$ is defined by $T(A,\lambda) = \{ x \in \Sigma^*, \lambda < p(x) \}$. If $x \in T(A,\lambda)$, we say that $x$ is accepted with cut-point $\lambda$. $T(A,\lambda)$ is called the set defined by $A$ with cut-point $\lambda$.

Definition 1.3.6. Let $U \subseteq \Sigma^*$. $U$ is a probabilistic cut-point event (PCE) if $U = T(A,\lambda)$ for some PA $A$ and some $\lambda \in [0,1)$.

For the case of any finite automaton $A$, $p(x) = 1$ if and only if $x \in T(A)$. If $x \notin T(A)$, then $p(x) = 0$. Hence, $T(A) = T(A,\lambda)$ for any $\lambda \in [0,1)$. Thus, every regular event is a PCE. Rabin constructed a particular probabilistic automaton $A$ to demonstrate that the converse is not true. For this $A$, $\{ T(A,\lambda) | \lambda \in [0,1) \}$ forms an uncountable pairwise different collection of sets. There are, however, only a countable number of regular events. Hence, this probabilistic automaton $A$ defines a set which is not defined by any finite automaton. Thus, the class of PCE's is a proper superset of $T$; that is, the class of probabilistic automata is more general than the class of finite automata. The class of PCE's is uncountable, but Paz showed that not every subset of $\Sigma^*$ is a PCE.

For any probabilistic automaton $A$, the state distribution function $\pi$ leads to a right invariant equivalence relation. We say, for any $x, y \in \Sigma^*$, $x \sim y$ iff $\pi(x) = \pi(y)$. Clearly, $\pi(xz) = \pi(yz) \forall z \in \Sigma^*$; thus $E$ is right invariant. $E$, in the case of probabilistic automata, may have up to a countable number of equivalence classes. However, in a manner similar to Theorem 1.2.1, any set of the form $T(A,\lambda)$ can be represented as the union of the equivalence classes of $E$ satisfying the condition that $\pi(x) \eta^p > \lambda$. 
Under certain conditions we can guarantee that there exists a finite automaton with the same acceptance set as a probabilistic automaton with a specific cut-point.

**Definition 1.3.7.** A cut-point \( \lambda \) is called isolated with respect to \( A \) if there exists a \( \delta > 0 \) such that \( |p(x) - \lambda| \geq \delta \) \( \forall x \in E^* \).

**Theorem 1.3.1.** Let \( A \) be a PA and \( \lambda \) an isolated cut-point for \( A \). Then there exists an PA \( B \) such that \( T(A, \lambda) = T(B) \). If \( A \) has \( n \) states, then \( B \) can be chosen to have \( n' \) states, where \( n' \leq (1 + \frac{1}{2\delta})^{n-1} \).

It seems that in passing from a probabilistic automaton to an equivalent finite automaton we may have to increase the number of states. In many cases a probabilistic automaton equivalent to a finite automaton can be constructed which is more economical in the state saving sense.

Paz found other conditions for the equivalence of probabilistic and finite automata. His analysis in this area also introduced other interesting mathematical results.

**Definition 1.3.8.** Two PA \( A \) and \( B \) are strongly equivalent iff \( T(A, \lambda) = T(B, \lambda) \) \( \forall \lambda \in [0, 1) \).

**Definition 1.3.9.** A probabilistic table (PT) is a system \( \theta = (S, M) \), where \( S \) and \( M \) are as in definition 1.3.2. A discriminative probabilistic table (DPT) is a system \( \tau = (S, M, F) \), where \( S \), \( M \), and \( F \) are as in definition 1.3.2.

**Definition 1.3.10.** Let \( T \) be a DPT and \( \pi_0 \) and \( \rho_0 \) any initial distributions for \( T \). A tape \( x \) distinguishes between \( \pi_0 \) and \( \rho_0 \) if
\( \pi_0^M(x) \eta^F \neq \rho_0^M(x) \eta^F \). Two initial distributions are k-equivalent if they are not distinguishable by any tape \( x \in \Sigma^* \) such that \( L(x) \leq k \).

These definitions lead immediately to results of equivalence of probabilistic automata.

**Theorem 1.3.3.** Let \( T \) be a DPT. If two initial distributions are distinguishable, then there exists \( x \in \Sigma^* \) such that \( L(x) \leq n-2 \) which distinguishes between them, where \( n \) is the cardinality of the state set \( S \).

**Theorem 1.3.4.** Given two PA \( A_1 \) and \( A_2 \) over \( \Sigma \), if \( p_1(x) = p_2(x) \ \forall x \in \Sigma^* \) such that \( L(x) \leq n_1 + n_2 - 2 \), then \( A_1 \) and \( A_2 \) are strongly equivalent.

Paz extended the concept of definiteness for finite automata to construct a body of results for automata equivalence and approximation. The \( k \)-suffix of any tape \( x \) of length at least \( k \) is the tape of the last \( k \) symbols of \( x \).

**Definition 1.3.11.** Let \( k \geq 0 \) and \( U \subseteq \Sigma^* \). \( U \) is a weakly \( k \)-definite set if for any \( x \in \Sigma^* \) such that \( L(x) \geq k \), then \( x \in U \) iff the \( k \)-suffix of \( x \) is in \( U \). \( U \) is \( k \)-definite if it is weakly \( k \)-definite but not weakly \((k-1)\)-definite. A set \( U \) is definite if it is \( k \)-definite for some \( k \).

Clearly, every definite set is definable by some finite automaton. Such finite automata are called definite automata. Paz sought to generalize this concept to form a class of definite probabilistic automata.
Definition 1.3.12. A PT $\theta = (S, M)$ is weakly $k$-definite iff, for any $x \in \mathcal{E}^*$ such that $L(x) \geq k$ and for any initial distributions $\pi_0$ and $\rho_0$, the following holds: $(\pi_0 - \rho_0)M(x) = 0$.

This, indeed, does generalize the notion of weak $k$-definiteness. Given any final set $F$ and initial distribution $\pi_0$ for a weakly $k$-definite probabilistic table, then for all $x \in \mathcal{E}^*$ such that $L(x) \geq k$, $p(x) = p(k$-suffix of $x)$.

Definition 1.3.13. A PT is $k$-definite if it is weakly $k$-definite but not weakly $(k-1)$ definite. A PT is definite if it is $k$-definite for some $k$.

Paz extended the result of Perles, Rabin, and Shamir [15] to include his probabilistic notion of definiteness.

Theorem 1.3.5. Let $\theta = (S, M)$ be a PT. If $\theta$ is definite, then it is $(n-1)$-definite, where $n$ is the cardinality of $S$.

Note that any probabilistic automaton with a definite probabilistic table defines a definite event and, hence, must have an equivalent definite automaton. Paz introduced an even weaker notion of definiteness, i.e., quasiddefiniteness, to define a larger class of automata. Members of this class do not, in general, have equivalent finite automata; however, this class possesses some interesting approximation properties.

Definition 1.3.14. Given any arbitrary $n$-dimensional vector $\xi$, we define $|\xi| = \max_i |\xi_i|$. Also, for any arbitrary $n \times n$ matrix $Q$ we define $|Q| = \max_{i,j} |q_{ij}|$. 


Definition 1.3.15. A PT $\theta = (S, M)$ is quasi-definite iff, for every $\varepsilon > 0$ there is a number $N(\varepsilon)$ such that for any $x \in \Sigma^*$ such that $L(x) \geq N(\varepsilon)$ and for any possible initial distributions $\pi_0$ and $\rho_0$, then $|\pi_0 - \rho_0| \cdot M(x) | < \varepsilon/n$.

Hence, for any probabilistic automaton $A$ with quasi-definite table let us fix $\varepsilon$; if $x \in \Sigma^*$ is any tape such that $L(x) \geq N(\varepsilon)$, then $|p(x) - p(N(\varepsilon)\text{-suffix of } x)| < \varepsilon$. Thus we have a looser, approximation notion of definiteness.

Definition 1.3.16. Let $M \in M_n$. $M$ is called scrambling if for every pair of states $i_1$ and $i_2$, there is a state $j$ such that $m_{i_1j} > 0$ and $m_{i_2j} > 0$. A finite subset $\{M_1, \ldots, M_k\}$ of $M_n$ satisfies the $H_4$ condition of order $r$ if there is an $r$ such that any product of the $M$'s of length at least $r$ is scrambling. For a PA, any tape $x \in \Sigma^*$ such that $M(x)$ is scrambling is called a scrambling tape.

Theorem 1.3.6. Let $\theta = (S, M)$ be a PT. $\theta$ is quasi-definite iff the set of matrices $\{M(\sigma) | \sigma \in \Sigma\}$ satisfies the $H_4$ condition.

Theorem 1.3.7. Let $\theta = (S, M)$ be any PT. There is an effective procedure which decides in a finite number of steps whether $\theta$ is quasi-definite.

Definition 1.3.17. Let $U \subseteq \Sigma^*$ and $U^c = \Sigma^* - U$. The set of tapes $U$ $\varepsilon$-approximates the set of tapes defined by a PA with cut-point $\lambda$ if $(T(\lambda, \lambda) - U) \cup (T(\lambda, \lambda)^c - U^c) \subseteq \{x | |p(x) - \lambda| < \varepsilon\}$. 
Theorem 1.3.8. Let $A$ be a PA with quasidfinite table. Then for every $\epsilon > 0$ there exists a definite automaton $A_{\epsilon}$ such that $T(A_{\epsilon})$ $\epsilon$-approximates $T(A, \lambda)$ for any given $\lambda$.

Theorem 1.3.9. If $A$ is a PA with quasidfinite table and if $\lambda$ is an isolated cut-point, then $T(A, \lambda)$ is a definite set.

Rabin defined the more specific concept of an actual automaton to derive a converse for Theorem 1.3.9. Note that every actual automaton has a quasidfinite table.

Definition 1.3.18. A PA $A$ is an actual automaton iff $m_{ij}(\sigma) > 0 \ \forall \sigma \in \Sigma \ \forall i,j=1,...,n$.

The use of actual automata is natural in simulating certain systems for which all of the transition probabilities are positive, no matter how small they are.

Theorem 1.3.10. Every definite set is definable by some actual automaton with isolated cut-point.

Paz also extended the theorem of Rabin concerning the effects of small perturbations of the transition probabilities of an automaton. Results in this realm are useful in considering the stability of properties of systems subject to perturbation and approximation.

Theorem 1.3.11. Let $A = (S,M,\pi_0,F)$ be a PA with quasidfinite table. Given any $\delta > 0$, there exists an $\epsilon = \epsilon(\delta,A) > 0$ such that for every PA $A_{\epsilon} = (S,M',\pi_0',F)$ with $|M(\sigma) - M'(\sigma)| < \epsilon \ \forall \sigma \in \Sigma$ and $|\pi_0 - \pi_0'| < \epsilon$, then $|p(x) - p_{\epsilon}(x)| < \delta \ \forall x \in \Sigma^*$. 
Corollary 1.3.12. Let $A$ and $A_{\epsilon}$ be PA's as in Theorem 1.3.11 and let $\lambda$ be an isolated cut-point for $A$. For any sufficiently small $\epsilon > 0$, $\lambda$ is an isolated cut-point for $A_{\epsilon}$ and $T(A, \lambda) = T(A_{\epsilon}, \lambda)$. 
CHAPTER II
Environments

2.1 The Concept of Environments for Automata

In discussing machines subject to error or probabilistic transition, we have considered those systems whose transition probabilities are only related to the present state and latest input symbol and are independent of any external factors. Now, however, we shall construct automata models to be more general and more realistic. The viewpoint taken is that every machine is located within an environment and the transition function of that machine is not only related to the input but also to the configuration of the environment.

We denote the set of possible values that the environment may assume by $E$. Presently, we do not make any stipulations as to the origin, form, structure, or cardinality of $E$. As each new input symbol is read by the machine, the configuration of the environment is observed and the transition ensues as a function of the input symbol and the condition of the environment. We also view the initial distribution as a function of the environment. Thus, we shall consider the total environment for a machine. The total environment is the sequence of environments that consists of the initial and the subsequent environments as each symbol is read.
Definition 2.1.1. Let $E^\infty$ denote the cartesian product of a countable number of copies of the environment set $E$. A total environment, or environment sequence, is a mapping $e(k) = e_k \in E$ for $k=0,1,2,\ldots$. We may also denote $e$ by the sequence $(e_0,e_1,e_2,\ldots)$.

2.2 Automata in Deterministic Environments

Here we shall consider the case in which the environment is completely specified. That is, for the initial distribution and subsequent transitions we are given the precise condition of the environment. In such a case we say that the environments for the initial distribution and subsequent transitions are obtained by a deterministic rule, the total environment.

Definition 2.2.1. An automaton in a deterministic environment (ADE) is a system $A = (\Sigma,S,\Sigma_0,F,E)$, where $\Sigma$ is a finite input alphabet, $S = \{s_1,\ldots,s_n\}$ is the finite state set, $\Sigma$ is a mapping $G: \Sigma \times E \to M_n$ (the basic matrix transition function), $\pi_0: E \to \Sigma_n$ (the initial distribution function), $F \subseteq S$ (the set of final states), and $E$ is the set of environments.

The matrix transition function $M$ is defined on $\Sigma^* \times E^\infty$ by the inductive recursion rule:

let $x \in \Sigma^*$ such that $L(x) = k$ and $e \in E^\infty$,
then $M(x\sigma,e) = M(x,e)G(\sigma,e_{k+1})$.

Hence, for all $x \in \Sigma^*$ and $e \in E^\infty$, $M(x,e) = \prod_{i=1}^{k} G(\sigma_i,e_i)$, where $x = \sigma_1 \ldots \sigma_k$ and $e = (e_0,e_1,\ldots)$. In considering the function $\pi_0$ we shall use the notation $\pi_0(e_0)$ and $\pi_0(e)$ interchangeably, where $e = (e_0,e_1,\ldots)$. 
As with probabilistic automata we are able to compute the state
distribution of $A$ after input $x$. Now, however, this will depend on
the total environment $e$. We denote the state distribution of $A$
after input $x$ in total environment $e$ by $\pi(x,e) = \pi_0(e_0)^H(x,e)$.
Similarly, we let $p(x,e) = \pi(x,e)^F$, where $\eta^F$ is as in definition
1.3.4, denote the probability that $A$ is in a state in $F$ after input
$x$ in total environment $e$.

**Definition 2.2.2.** Let $A$ be an ADE and $\lambda$ a real number,
$\lambda \in [0,1)$. The set of tapes $T(A,e,\lambda)$ defined by

$$T(A,e,\lambda) = \{x | x \in \Sigma^*, \lambda < p(x,e)\}$$

is called the set of tapes accepted, or defined, by $A$ in total environ-
ment $e$ with cut-point $\lambda$.

**Definition 2.2.3.** Let $Q(e_0,\ldots,e_k) = \{e \in \Sigma^* | e(i) = e_i \text{ for } i=0,1,\ldots,k\}$
and $Q'(e_1,\ldots,e_k) = \{e \in \Sigma^* | e(i) = e_i \text{ for } i=1,2,\ldots,k\}$.

**Lemma 2.2.1.** Let $A$ be an ADE. For any $k \geq 0$ and any fixed
$e_0,e_1,\ldots,e_k \in \Sigma^*$, the following functions are constant in $Q(e_0,\ldots,e_k)$:
$M(x,\cdot)$, $\pi(x,\cdot)$, and $p(x,\cdot)$ for each $x \in \Sigma^*$ such that $L(x) \leq k$.

**Proof:** Let $k \geq 0$; fix $e_0,e_1,\ldots,e_k \in \Sigma^*$ and fix $x = \sigma_1,\ldots,\sigma_r$
such that $L(x) = r \leq k$. Let $e' \in Q(e_0,\ldots,e_k)$. Then $e' =$
$(e_0',e_1',\ldots,e_k',e_{k+1}',e_{k+2}',\ldots)$, where $e_{k+1}',e_{k+2}',\ldots \in \Sigma^*$. So
$M(x,e') = \prod_{i=1}^{r} G(\sigma_i,e_i)$. Thus, $M(x,\cdot)$ is constant in $Q(e_0,\ldots,e_k)$. Simi-
larly, $\pi(x,e') = \pi_0(e_0) \prod_{i=1}^{r} G(\sigma_i,e_i)$ and $p(x,e') = \pi_0(e_0) \prod_{i=1}^{r} G(\sigma_i,e_i)^F$
are also constant in $Q(e_0,\ldots,e_k)$. Note that $M(x,\cdot)$ is also constant
in $Q'(e_1,\ldots,e_k)$.

$\square$
2.3 Mathematical Generality of the ADE Model

Let \( A' = (S', M', \pi_0', F') \) be any probabilistic automaton over the alphabet \( \Sigma' \). Let \( \Sigma \) be any nonempty set; let us fix \( \epsilon \in \Sigma \). We can now construct an ADE \( A = (\Sigma, S, G, \pi_0, F, \epsilon) \) such that \( T(A', \lambda) = T(A, (e, e, \ldots, \lambda)) \) for all \( \lambda \in \{0, 1\} \). We simply let \( \Sigma = \Sigma', \quad S = S', \quad F = F', \quad \pi_0(e) = \pi_0', \quad G(\sigma, e) = M'(\sigma) \text{ for } \sigma \in \Sigma \). Thus, \( p(x, e) = \pi_0(e)M(x, (e, e, \ldots))\eta^F = \pi_0 \left( \prod_{i=1}^{k} G(\sigma_i, e) \right) \eta^F = \pi_0 \left( \prod_{i=1}^{k} M'(\sigma_i) \right) \eta^F = p'(x) \). Therefore, for any \( \lambda \in \{0, 1\} \), \( x \in T(A', \lambda) \) iff \( x \in T(A, (e, e, \ldots, \lambda)) \) and, hence, every PCE can be defined by an ADE with some total environment \( \epsilon \in \Sigma^\infty \).

Let \( \Sigma = \{\sigma\} \). Paz [14] demonstrated that there exists \( U \subset \Sigma^* \) such that \( U \) is not a PCE. Let us now consider a two-state ADE \( A \) over \( \Sigma^* \) where \( E = \{0, 1\} \), \( \pi_0(e) = (1-e, e) \), \( G(\sigma, e) = \begin{bmatrix} 1-e & e \\ 1-e & e \end{bmatrix} \) \text{ for } \sigma \in \Sigma \), and \( \eta^F = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \). Let us denote a tape of \( k \) repetitions of the symbol \( \sigma \) by \( \sigma^k = \sigma \sigma \ldots \sigma \). Note that \( \sigma^0 = \lambda \), the empty tape. Thus, for our particular situation all of the elements of \( \Sigma^* \) are of the form \( \sigma^k \) for some \( k \geq 0 \). Let \( U \) be any subset of \( \Sigma^* \); then \( U = \{\sigma^{k_1}, \sigma^{k_2}, \ldots\} \). We construct \( e(k) = \begin{cases} 1 & \text{for some } i \geq 0 \\ 0 & \text{elsewhere} \end{cases} \)

Such an \( e \) is an element of \( \Sigma^\infty \). Now, for \( k \geq 1 \) we have

\[
M(\sigma^k, e) = \begin{cases} 
\begin{bmatrix}
0 & 1 \\
0 & 1
\end{bmatrix} & \text{if } k = k_i \text{ for some } i \\
\begin{bmatrix}
1 & 0 \\
1 & 0
\end{bmatrix} & \text{elsewhere}
\end{cases}.
\]

Therefore, \( p(\sigma^k, e) = \pi_0(e)M(\sigma^k, e)\eta^F = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \) if \( k = k_i \) for some \( i \)

When \( k = 0 \), \( p(\sigma^0, e) = p(\lambda, e) = \pi_0(e)\eta^F = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \) if \( k_i = 0 \) for some \( i \)

Let \( \lambda = 0 \), then \( U = T(\lambda, e, \lambda) \). Since \( U \) is arbitrary, every subset of our particular \( \Sigma^* \) is definable by \( A \) under some environment. But
there exists \( U \subseteq \Sigma^* \) which is not definable by a probabilistic automaton. Thus, the class of automata in deterministic environments properly contains the class of probabilistic automata.

Let \( A \) be an ADE and let \( e \in \Sigma^\infty \). For any \( x, y \in \Sigma^* \) such that \( \pi(x, e) = \pi(y, e) \), then \( \pi(xz, e) = \pi(yz, e) \) for all \( z \in \Sigma^* \). Hence, for any \( e \in \Sigma^\infty \), the state distribution function \( \pi(\cdot, e) \) is a right invariant equivalence relation on \( \Sigma^* \). As with probabilistic automata, for any fixed \( e \in \Sigma^\infty \) there may be up to a countable number of equivalence classes of the relation \( \pi(\cdot, e) \). For any \( \lambda \in [0,1) \), \( T(A, \mathbf{e}, \lambda) \) is the union of the equivalence classes satisfying the condition that \( \pi(x, \mathbf{e}) \eta^F > \lambda \).

The greater generality of the ADE model manifests itself in the fact that the equivalence relation and, hence, the equivalence classes may depend on \( \mathbf{e} \). Depending on \( A \), there may be an uncountable number of unique equivalence relations. The above two state ADE has an uncountable number of unique equivalence relations.

2.4 Finite Environment Set

Let \( A \) be an ADE with a finite environment set. We shall construct an effective procedure to find a probabilistic automaton which simulates the operation of \( A \).

**Definition 2.4.1.** A PA \( A' \) over alphabet \( \Sigma' \) simulates the ADE \( A = (\Sigma, S, G, \pi_0, F, E) \) if there is a relation \( R \) between \( \Sigma'^* \) and \( \Sigma^* \times \Sigma^\infty \) such that

1) for any \( (x, e) \in \Sigma^* \times \Sigma^\infty \) there exists \( x' \in \Sigma'^* \) such that \([x', (x, e)] \in R\),
2) if \([x', (x, e)] \in R\), then for each \(\lambda \in [0,1]\), \(x' \in T(A', \lambda)\) if and only if \(x \in T(A, \lambda)\);

3) if for some \(\lambda \in [0,1]\), \(x' \in T(A', \lambda)\), then there exists \((x, e) \in \Sigma^* \times E^\omega\) such that \([x', (x, e)] \in R\).

If a PA \(A'\) simulates an ADE \(A\), given any input \(x\) for \(A\) under total environment \(e\), we have a rule \(R\) to find an input tape \(x'\) for \(A'\) such that \(x'\) has the same acceptance properties as \(x\) for \(A\) under total environment \(e\). A tape \(x' \in \Sigma^*\) is admissible if and only if there is an input tape \(x\) for \(A\) under some total environment \(e\) from which the rule \(R\) will yield \(x'\). By condition 3 of definition 2.4.1 we see that tapes which are not admissible are not accepted by \(A'\) for any \(\lambda \in [0,1]\), that is, if \(x'\) is not admissible, then the probability of acceptance of \(x'\) by \(A'\) is zero. By lemma 2.2.1 we see that the probability of acceptance of any tape \(x \in \Sigma^*\) depends only on the environments \(e_0, e_1, \ldots, e_k\), where \(k = L(x)\). Hence, the relation \(R\) need only depend on \(x\) and \(e_0, e_1, \ldots, e_k\), where \(k = L(x)\). Thus, all pairs of the form \((x, e)\), where \(e \in Q(e_0, \ldots, e_k)\) and \(k = L(x)\), can be considered equivalent.

**Theorem 2.4.1.** Let \(A = (\Sigma, S, G, \pi_0, \mathcal{F}, E)\) be an ADE such that \(\#(E) < \infty\). There exists a PA \(A'\) over some finite alphabet \(\Sigma'\) which simulates \(A\).

**Proof:** Let \(m = \#(\Sigma)\) and \(E = \#(E)\). Without loss of generality we shall let \(\Sigma = \{1, 2, \ldots, m\}\) and \(E = \{0, 1, \ldots, E-1\}\). Let us consider \(\Sigma' = \{0, 1, \ldots, E(m+1)-1\}\). Let \(\phi\) be any symbol such that \(\phi \notin \Sigma\). We now define a mapping \(g: (\Sigma \cup \{\phi\}) \times E \to \Sigma'\) as follows:

---

\(^1\) \#(E) denotes the cardinality of the set \(E\).
\[ g(\sigma, e) = e + \sigma E \quad \forall \sigma \in \Sigma \text{ and } \forall e \in E \]
\[ g(\phi, e) = e \quad \forall e \in E. \]

This mapping is a one-to-one correspondence. We extend \( g \) to \(((\Sigma u\{\phi\}) \times E)^*\), the set of all finite sequences of symbols of \((\Sigma u\{\phi\}) \times E\), by component-wise application of the mapping and concatenation of the results. That is, \( a_1, a_2, \ldots, a_k \in (\Sigma u\{\phi\}) \times E \), then 
\[ a_1 a_2 \ldots a_k \in ((\Sigma u\{\phi\}) \times E)^* \quad \text{and} \quad g(a_1 a_2 \ldots a_k) = g(a_1) g(a_2) \ldots g(a_k). \]

\( g \) remains a one-to-one correspondence between \(((\Sigma u\{\phi\}) \times E)^*\) and \( \Sigma^* \times E^\omega \).

Let \( b \) be any nonempty element of \(((\Sigma u\{\phi\}) \times E)^*\), then \( b \) is isomorphic to the form \((y, (e_0, \ldots, e_k))\), where for some \( k \geq 0 \) \( y \in (\Sigma u\{\phi\})^* \), \( L(y) = k+1 \), and \( e_0, \ldots, e_k \in E \). We define the relation \( F \) between \( \Sigma^* \times E^\omega \) to be the set of all elements of the form: \([g(\phi x, (e_0, \ldots, e_k)), (x, e)]\), where \( x \in E^* \), \( L(x) = k \), and \( e = (e_0, \ldots, e_k, \ldots) \). Note that when \( x = \lambda \in E^* \), then \( L(\lambda) = 0 \) and \( \phi \lambda = \phi \).

We shall now construct a PA \( A' \) to simulate \( A \). Let 
\[ A' = (S', M', \pi_0', F') \] be a PA over \( E' \). Let \( S' = Su\{s_{n+1}, s_{n+2}\}, \) where \( S = \{s_1, \ldots, s_n\} \) is the state set of \( A \). For any \( \sigma' \in E \subset E' \), then \( 0 \leq \sigma' \leq E'-1 \). So for any \( \sigma' \) such that \( 0 \leq \sigma' \leq E'-1 \), we define 
\[
M'(\sigma') = \begin{bmatrix}
\ldots & 0 & 1 \\
\ldots & \ldots & \vdots \\
\ldots & \ldots & \ldots & 0 & 1 \\
\pi_0'(1)(\sigma') & \ldots & \pi_0'(n)(\sigma') & 0 & 0 \\
0 & \ldots & 0 & 0 & 1
\end{bmatrix}
\]

where \( \pi_0'(1)(\sigma') \) is the \( i \)-th component of \( \pi_0(\sigma') \). Note that \( 0 \leq \sigma' \leq E'-1 \) implies \( \sigma' \in E \). These values of \( \sigma' \) are used to incorporate the initial distribution function of \( A \) into the PA model \( A' \). The values
\( \sigma' \in \Sigma^i \) such that \( E \leq \sigma \leq E(m+1) - 1 \) are used in order that \( A' \) can imitate the basic matrix transition function of \( A \). The set \( \{ \sigma' | E \leq \sigma' \leq E(m+1) - 1 \} \) is in one-to-one correspondence with the set \( \Sigma \times E \) under the mapping \( g \). So for \( E \leq \sigma' \leq E(m+1) - 1 \), define

\[
M'(\sigma') = \begin{pmatrix}
G(g^{-1}(\sigma')) \\
\vdots \ & \vdots \\
0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0 & 0 & 1
\end{pmatrix}
\]

where \( G \) is the basic matrix transition function of \( A \). Also, let \( \pi^i_0 = (0, \ldots, 0, 1, 0) \) and \( F' = F \). Therefore \( \eta^F = \begin{pmatrix} \eta^F \\ 0 \\ 0 \end{pmatrix} \).

It is clear from the definition of \( R \) that for any \( (x, e) \in \Sigma \times E^\infty \) there exists a unique \( x' = g(\phi x, (e_0, \ldots, e_k)) \), where \( k = L(x) \), such that \( [x', (x, e)] \in R \).

Suppose \( [x', (x, e)] \in R \). Therefore, \( x' = g(\phi, e_0) = e_0 \). Let \( p'(x') \) be the probability that \( A' \) accepts \( x' \) and \( p(x, e) \) be the probability that \( A \) accepts \( x \) under total environment \( e \). Hence,

\[
p'(x') = p'(e_0) = \pi_0^i M'(e_0) \eta^F
\]

\[
= (0, \ldots, 0, 1, 0) \begin{pmatrix}
0 & \cdots & \cdots & 0 \\
\vdots \ & \vdots \\
\pi_0(1) (e_0) & \cdots & \pi_0(n) (e_0) & 0 & 0 \\
0 & \cdots & \cdots & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix} \eta^F \\ 0 \\ 0 \end{pmatrix}
\]

\[
= \pi_0(e_0) \eta^F = \pi_0(e_0) M(A, e) \eta^F = p(A, e).
\]

Thus, for each \( \lambda \in [0, 1) \), \( p'(x') = p'(e_0) > \lambda \) iff \( p(A, e) > \lambda \). That is, if \( [x', (A, e)] \in R \), then for each \( \lambda \in [0, 1) \), \( x'^T(A', \lambda) \) iff \( A \in T(A, e, \lambda) \).
Suppose \([x', (x, e)] \in R\), where \(x \neq \Lambda\). Hence, \(L(x) > 0\). By the definition of \(R\) we have that \(x' = g(\phi x, (e_0, \ldots, e_k))\), where \(x = \sigma_1, \ldots, \sigma_k\) has length \(k\). So \(x' = g(\phi, e_0)g(\sigma_1, e_1), \ldots, g(\sigma_k, e_k) = e_0\sigma'_1, \ldots, \sigma'_k\). Thus, \(p'(x') = p'(e_0\sigma'_1, \ldots, \sigma'_k) = \pi'_0 M'(e_0\sigma'_1 \ldots \sigma'_k) n^{E'}\). But

\[
\prod_{i=1}^{k} M'(\sigma'_i) = \prod_{i=1}^{k} \begin{bmatrix}
G(g^{-1}(\sigma'_i)) & 0 & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 \\
0 & \ldots & 0
\end{bmatrix}
\]

\[
= \prod_{i=1}^{k} \begin{bmatrix}
G(\sigma_i, e_i) & 0 & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 \\
0 & \ldots & 0
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\prod_{i=1}^{k} G(\sigma_i, e_i) & 0 & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 \\
0 & \ldots & 0
\end{bmatrix}
\]

\[
= \begin{bmatrix}
M(x, e) & 0 & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 \\
0 & \ldots & 0
\end{bmatrix}
\]

Hence, we have
\[ p'(x') = \pi_0^t M'(e_0) \left[ \begin{array}{c|c|c}
M(x,e) & 0 & 0 \\
\vdots & \vdots & \vdots \\
0 & 0 & 0 \\
0.0 & 0.0 & 0.0 \\
\end{array} \right] F' \]

\[ = (0, \ldots, 0, 1, 0) \left[ \begin{array}{c|c|c}
0 & 1 & 0 \\
\vdots & \vdots & \vdots \\
0 & 1 & 0 \\
\pi_0^t(e_0) & \pi_0^n(e_0) & 0 \\
0 & 0 & 0 \\
\end{array} \right] \left[ \begin{array}{c|c|c}
M(x,e) & 0 & 0 \\
\vdots & \vdots & \vdots \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{array} \right] F' \]

\[ = \pi_0^t(e_0)M(x,e)F' = p(x,e). \]

So, if \([x', (x,e)] \in R\) and \(x \neq A\), then for each \(\lambda \in [0,1)\) \(p'(x') > \lambda\) iff \(p(x,e) > \lambda\). Thus, \(x' \in T(A', \lambda)\) iff \(x \in T(A, e, \lambda)\).

We have verified that \(\forall x \in \Sigma^*\) and \(\forall e \in E^\infty\) if \([x', (x,e)] \in R\), then for each \(\lambda \in [0,1)\), \(x' \in T(A', \lambda)\) iff \(x \in T(A, e, \lambda)\).

Let \(x' \in \Sigma^*\) such that there does not exist \(x \in \Sigma^*\) and \(e \in E^\infty\) so that \([x', (x,e)] \in R\). Hence, there does not exist \(x \in \Sigma^*\) and \(e \in E^\infty\) such that \(x' = g(x, (e_0, \ldots, e_k))\), where \(k = L(x)\). Therefore, such an \(x'\) is not admissible. However, \(x' = g(b)\) for some \(b \in ((\Sigma U \phi) \times E)^*\) since \(g\) is a one-to-one correspondence between \(((\Sigma U \phi) \times E)^*\) and \(\Sigma^*\).

\(A',\) the empty tape for \(A'\), is the image of \(A^0\), the empty tape in \(((\Sigma U \phi) \times E)^*\), under \(g\). Clearly, \(A'\) is not admissible. \(p'(A') = \pi_0^tF' = (0, \ldots, 0, 1, 0) 0 = 0\). Hence \(A' \notin T(A', \lambda)\) for any \(\lambda \in [0,1)\).

Let \(x' \in \Sigma^*\) be any tape which is not admissible and \(x' \neq A'\).
\(x' = \sigma_0^t \ldots \sigma_k^t\) is the image under \(g\) of some element \(b_0((\Sigma U \phi) \times E)^*\),
where \( h \) has the form \((y,(e_0,\ldots,e_k))\) and \( L(x') = L(y) = k+1 \).

Note that \( y \epsilon (\Sigma \cup \phi) \). Let \( y = \tau_0 \tau_1 \ldots \tau_k \), where \( \tau_i \epsilon \Sigma \cup \phi \) for \( i=0,1,\ldots,k \). \( x' \) is not admissible iff \( \tau_0 \neq \phi \) or \( \tau_i \epsilon \Sigma \) for some \( i=1,2,\ldots,k \). If \( \tau_0 \neq \phi \), then \( \tau_0 \epsilon \Sigma \). \( \sigma_0' = g(\tau_0,e_0) \) and, hence \( E \leq \sigma_0' \leq E(m+1)-1 \). By our definition of \( M' \) we see that \( A' \) under input \( \sigma_0' \sigma_1' \ldots \sigma_k' \) enters state \( s_{n+2} \) at the first transition. But \( s_{n+2} \) is an absorbing state and \( s_{n+2} \not\epsilon \mathcal{F}' \). Hence, \( p'(x') = 0 \).

So \( x' \epsilon \mathcal{T}(A',\lambda) \) for any \( \lambda \epsilon [0,1) \). If \( \tau_i \epsilon \Sigma \) for some \( i=1,2,\ldots,k \); say \( \tau_j \epsilon \Sigma \). Then \( \tau_j = \phi \). \( \sigma_j' = g(\tau_j,e_j) \) and, hence, \( 0 \leq \sigma_j' \leq E-1 \). By our definition of \( M' \) we see that \( A' \) under input \( \sigma_0' \sigma_1' \ldots \sigma_k' \) enters state \( s_{n+2} \) at the \((j+1)\)th transition. Again, \( s_{n+2} \) is an absorbing state and \( s_{n+2} \not\epsilon \mathcal{F}' \). Hence, \( p'(x') = 0 \) and \( x' \epsilon \mathcal{T}(A',\lambda) \) for any \( \lambda \epsilon [0,1) \).

Thus, if for some \( \lambda \epsilon [0,1) \) \( x' \epsilon \mathcal{T}(A',\lambda) \), then \( x' \) is admissible; that is, there exists \( (x,e) \epsilon \mathcal{E}^* \times E^\infty \) such that \( [x',(x,e)] \epsilon R \).

Consequently, \( A' \), as constructed, simulates \( A \). \( \square \)

The results of section 2.3 and theorem 2.4.1 seem to present a paradox. We have found for a particular \( \Sigma^* \) that there exists \( U \epsilon \Sigma^* \) which is not a PCE; that is, for any \( \lambda \epsilon [0,1) \) there is no probabilistic automaton \( \hat{A}_p \) such that \( U = \mathcal{T}(\hat{A}_p,\lambda) \). However, we showed that there exists an ADE with finite environment set over \( \Sigma \) such that \( U = \mathcal{T}(\hat{A}_e,\lambda) \) for some \( e \epsilon E^\infty \) and \( \lambda \epsilon [0,1) \). By theorem 2.4.1 there exists a probabilistic automaton \( A' \) over an expanded alphabet \( \Sigma' \) such that \( A' \) simulates \( \hat{A} \). Let \( U' = \{ g(\phi x,e) | x \epsilon U \} \), where \( e \) is the total environment whereby \( U = \mathcal{T}(\hat{A}_e,\lambda) \). \( U' = \mathcal{T}(A',\lambda) \) and, hence, is a PCE. The paradox is resolved by noting that we not only have expanded \( \Sigma \) to \( \Sigma' \)
but also enriched the alphabet with a structure that had been included
in the environment. The problem for \( U \subseteq \Sigma^* \) remains because \( U' \) has a
different structure within \( \Sigma'^* \). In fact, we can find a subset of \( \Sigma'^* \)
which is not a PCE.

**Example 2.4.1.** As an example of the application of theorem 2.4.1,
let us consider the ADE discussed in section 2.3. \( A \) is a two-state
ADE over \( \Sigma = \{\sigma\} \) with \( \Sigma = \{0,1\} \). \( \pi_0(e) = (1-e,e), G(\sigma,e) = (1-e,e) \),
and \( \eta^F = (0,1) \). By theorem 2.4.1 \( A' \) has alphabet \( \Sigma' = \{0,1,2,3\} \)
and \( M' \) is defined on \( \Sigma' \) as follows

\[
M'(0) = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\qquad
M'(1) = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

\[
M'(2) = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\qquad
M'(3) = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

The mapping \( g \) is described by:

\[ g(\phi,e) = e \quad \text{and} \quad g(\sigma,e) = 2+e . \]

With \( \pi'_0 = (0,0,1,0) \) and \( \eta^F = (1,0) \), it is easily verified by the same
procedure as the proof that for any \( \lambda \in [0,1] \) \( T(A',\lambda) = \{ g(\phi x,e) | e \in E^\infty \}
\]
and \( x \in T(A,\varepsilon,\lambda) \) and that \( A' \) simulates \( A \).

**Corollary 2.4.2.** Let \( A = (E, S, G, \pi_0, F, E) \) be an ADE. For any
fixed \( e \in E^\infty \) such that \( E = \{ e_i | i=0,1,\ldots \} \) is a finite set, then there
exists a PA \( A' \) over some alphabet \( \Sigma' \) such that for any \( x \in \Sigma^* \)
and $\lambda \in \{0,1\}$, $x \in T(A,e,\lambda)$ iff $g_E(x,e) \in T(A',\lambda)$, where $g_E$ is defined for the ADE $A = (\Sigma,S,G,\pi_0,F,E)$ as the function $g$ in theorem 2.4.1.

2.5 Reducing the Environment Set and the Induced Metric

Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE and let $e$ and $e'$ be two distinct configurations of the environment. If $\pi_0(e) = \pi_0(e')$ and $G(\sigma,e) = G(\sigma,e')$ for all $\sigma \in \Sigma$, then $e$ and $e'$ have identical effects on the transition functions. That is, the configurations $e$ and $e'$ are equivalent in the sense that for any $x \in \Sigma^*$ and $e \in E^\omega$, $\pi(x,e)$ remains constant if we substitute $e$ for $e'$ whenever $e'$ occurs in $e$. Thus, we do not need to include $e'$ in the environment set since it does not augment the structure of $A$. Moreover, we can reduce $E$ so that no two distinct environments are equivalent in the above sense.

**Definition 2.5.1.** Let $A = (\Sigma,S,G,\pi_0,F,E)$ be an ADE, then we say $E$ is irreducible if for all $e, e' \in E$, $e = e'$ iff $\pi_0(e) = \pi_0(e')$ and $G(\sigma,e) = G(\sigma,e')$ for all $\sigma \in \Sigma$.

Let us assume now that $E$ is an irreducible environment set for the ADE $A = (\Sigma,S,G,\pi_0,F,E)$. Recalling definition 1.3.14, we construct a metric on $E$. Consider the function $d(e,e') = \max(|\pi_0(e) - \pi_0(e')|, \max_{\sigma \in \Sigma} |G(\sigma,e) - G(\sigma,e')|)$, where $e, e' \in E$. Clearly, $d(e,e') = d(e',e)$ and $d(e,e') \geq 0$ for all $e, e' \in E$. For any $e, e', e'' \in E$, then

$$d(e,e') = \max(|\pi_0(e) - \pi_0(e')|, \max_{\sigma \in \Sigma} |G(\sigma,e) - G(\sigma,e')|)$$

$$\leq \max(|\pi_0(e) - \pi_0(e'')| + |\pi_0(e'') - \pi_0(e')|),$$

$$\max_{\sigma \in \Sigma} (|G(\sigma,e) - G(\sigma,e'')| + |G(\sigma,e'') - G(\sigma,e)|)$$

$$\leq \max_{\sigma \in \Sigma} (|\pi_0(e) - \pi_0(e'')|, \max_{\sigma \in \Sigma} |G(\sigma,e) - G(\sigma,e'')|)$$
\[ d(e, e') = \max(\pi_0(e') - \pi_0(e''), \max_{\sigma \in \Sigma} |G(\sigma, e') - G(\sigma, e'')|) \]

So \( d \) satisfies the triangle property. Also, if \( e = e' \), then 
\( \pi_0(e) = \pi_0(e') \) and \( G(\sigma, e) = G(\sigma, e') \) for all \( \sigma \in \Sigma \). So \( d(e, e') = 0 \) when \( e = e' \). Since \( E \) is irreducible, if \( e \neq e' \), then by definition \( d(e, e') > 0 \). So the function \( d \) as defined above is a metric on \( E \).

**Theorem 2.5.1.** If \( E \) is an irreducible environment for an ADE \( A \), then the function \( d(e, e') = \max(\pi_0(e) - \pi_0(e'), \max_{\sigma \in \Sigma} |G(\sigma, e) - G(\sigma, e')|) \) is defined for all \( e, e' \in E \) and is a metric on \( E \).

Since irreducibility is needed only to show that if \( e \neq e' \), then \( d(e, e') \neq 0 \), the function \( d \) is a pseudo-metric on any arbitrary environment set.

Let us now consider an arbitrary ADE \( A = (\Sigma, S, G, \pi_0, F, E) \) for which the environment set is reducible to a finite set \( E^t \). That is, we can define an onto mapping \( h: E \rightarrow E^t \subset E \), where if \( h(e) = e^t \), then \( d(e, e^t) = 0 \). The mapping \( h \) is unique only when \( E^t \) is irreducible. \( h \) can be extended to \( h: E^\infty \rightarrow (E^t)^\infty \) by component-wise application of \( h \). That is, \( h(e) = (h(e_0), h(e_1), \ldots) \). So for all \( e \in E \), \( d(e, h(e)) = 0 \) and, hence, we know that \( \pi_0(e) = \pi_0(h(e)) \) and \( G(\sigma, e) = G(\sigma, h(e)) \) for all \( \sigma \in \Sigma \). So by extension, for all \( e \in E^\infty \), \( M(x, e) = M(x, h(e)) \), \( \pi(x, e) = \pi(x, h(e)) \), and \( p(x, e) = p(x, h(e)) \) for all \( x \in E^* \).

**Theorem 2.5.2.** Let \( A = (\Sigma, S, G, \pi_0, F, E) \) be any ADE, then \( T(A, e, \lambda) = T(A, h(e), \lambda) \) for all \( e \in E^\infty \) and \( \lambda \in (0, 1) \). Furthermore, if \( \#(h(E)) < \infty \), then \( A \) can be simulated by a PA.
Proof: Let \( e \in E^\infty \). We have seen that \( p(x,e) = p(x,h(e)) \) for all \( x \in E^* \). So for any \( \lambda \in [0,1) \), \( p(x,e) > \lambda \) iff \( p(x,h(e)) > \lambda \). Hence, \( x \in T(A,e,\lambda) \) iff \( x \in T(A,h(e),\lambda) \); that is, \( T(A,e,\lambda) = T(A,h(e),\lambda) \).

Let \( E^+ = h(E) \). So if \( \#(E^+) < \infty \), then by theorem 2.4.1 there exists a PA \( A' \) over alphabet \( \Sigma' \) which simulates the ADE \( A^+ = (\Sigma,S,G,\pi_0,F,E^+) \). Thus, there exists a relation \( R^+ \) between \( \Sigma'^* \) and \((\Sigma^+u\{\phi\})xE^+\) satisfying the conditions of definition 2.4. We now define a relation \( R \) between \( \Sigma'^* \) and \((\Sigma^+u\phi)xE^\infty\) by the rule:

\[
[x',(x,e)] \in R \quad \text{iff} \quad [x',(x,h(e))] \in R^+.
\]

Clearly, for all \( (x,e) \in \Sigma'^* \times E^\infty \), there exists \( x' \in \Sigma'^* \), such that \( [x',(x,e)] \in R \). If \( [x',(x,e)] \in R \), then \( [x',(x,h(e))] \in R^+ \). Hence, for any \( \lambda \in [0,1) \), \( x' \in T(A',\lambda) \) iff \( x \in T(A^+e,h(e),\lambda) \).

But \( T(A,e,\lambda) = T(A^+e,h(e),\lambda) \) for all \( \lambda \in [0,1) \). Therefore, if \( [x',(x,e)] \in R \), then for each \( \lambda \in [0,1) \), \( x' \in T(A',\lambda) \) iff \( x \in T(A^+e,h(e),\lambda) \). If for any \( x' \in \Sigma'^* \) there does not exist \( (x,e) \in \Sigma'^* \times E^\infty \) such that \( [x',(x,e)] \in R \), then since \( h \) is an onto mapping, there does not exist \( (x,e^+) \in \Sigma'^* \times (E^+) \) such that \( [x',(x,e^+)] \in R^+ \), and conversely. That is, \( x' \) is not admissible under the relation \( R \) iff \( x' \) is not admissible under the relation \( R^+ \). Hence, \( A' \) simulates \( A \) via the relation \( R \) as defined.

\[ \square \]

Corollary 2.5.3. Let \( A = (\Sigma,S,G,\pi_0,F,E) \) be an ADE. For any fixed \( e \in E^\infty \) such that \( E = \{e_i| i=0,1,\ldots\} \) is reducible to a finite set, then there exists a PA \( A' \) over some alphabet \( \Sigma' \) such that for any \( x \in \Sigma'^* \) and \( \lambda \in [0,1) \), \( x \in T(A,e,\lambda) \) iff \( g(x,h(e)) \in T(A',\lambda) \), where \( g \) is defined in theorem 2.4.1 and \( h \) is defined in theorem 2.5.2.
2.6 The Environment Set as a Semi-group

Definition 2.6.1. The pair \((W, \circ)\), where \(W\) is any set and \(\circ\) is any binary operation, is a semi-group if \(W\) is closed under the operation \(\circ\) and \(\circ\) is associative.

The set \(M_n^\Sigma\) of all \(n \times n\) stochastic matrices and the multiplication operation forms a semi-group. In Section 1.2 we observed that any \(\Sigma^*\) with the operation of concatenation forms a semi-group. Let \(A\) be any PA and \(x, y \in \Sigma^*\), then \(M(xy) = M(x)M(y)\). Since \(M_n^\Sigma\) is associative under the operation of multiplication, \(A\) defines a subsemi-group \(\{M(x) \mid x \in \Sigma^*\}\). We shall now generalize this result.

Lemma 2.6.1. Let \((W, \circ)\) be any semi-group. \(\{M(w) \in M_n^\Sigma \mid w \in W\}\) is a semi-group iff the mapping \(h: w \mapsto M(w)\) is a homomorphism.

Proof: Obvious from the definition of homomorphism and the fact that \(M_n^\Sigma\) is a semi-group.

Suppose we have an ADE \(A = (\Sigma, S, G, \pi_0, F, E)\), where \(E\) and some binary operation \(\circ\) form a semi-group. For each \(\sigma \in \Sigma\) let the mapping \(G(\sigma, \cdot) : E \rightarrow M_n^\Sigma\) be a homomorphism. By lemma 2.6.1 for each \(\sigma \in \Sigma\) the set \(G(\sigma) = \{G(\sigma, e) \mid e \in E\}\) is a semi-group under the operation of multiplication. That is, for any \(e_1, e_2 \in E\), then \(e_1 \circ e_2 \in E\) and \(G(\sigma, e_1 \circ e_2) = G(\sigma, e_1)G(\sigma, e_2)\). Also, since \(G(\sigma) \subset M_n^\Sigma\), multiplication of elements of \(G(\sigma)\) is associative. Under certain conditions an ADE on an environment set which is a semi-group may be simulated by a probabilistic automaton.
Theorem 2.6.1. Let $A = (E, S, G, \pi_0, F, E)$ be an ADE such that:

1. $E = S(E^+)$, where $E^+$ is any finite set and $S(E^+)$ denotes the semi-group generated by $E^+$ and the binary operation $\circ$;
2. for any $\sigma \in E$ the mapping $G_\sigma : E \rightarrow M_n$, defined as $G_\sigma(e) = G(\sigma, e)$, is a homomorphism;
3. $\pi_0(\cdot)$ is any fixed constant vector in $V_n$ for all $e \in E$.

Then there exists a PA $A'$ which simulates $A$.

Proof: $E = S(E^+)$ implies $E^+ \subseteq E$. Also, any $e \in E$ can be expressed as a finite product of elements of $E^+$. That is, $e = e_1^+ \cdots e_k^+$, where $e_i^+ \in E^+$, $i = 1, \ldots, k$ and $k$ is finite. Since for any $\sigma \in E$ $G_\sigma$ is a homomorphism, then $G(\sigma, e) = \prod_{i=1}^k G(\sigma, e_i^+)$. Clearly, if $e = e_1^+ \circ \cdots \circ e_k^+$, where $e_i^+ \in E^+$, $i = 1, \ldots, k'$ and $k'$ is finite, is any other decomposition of $e$, then $\prod_{i=1}^k G(\sigma, e_i^+) = \prod_{i=1}^{k'} G(\sigma, e_i^+)$. Let $x \in E^*$ and $e \in E^\infty$. Suppose $L(x) = r$. For each $e_i \in E^+$, $i = 1, \ldots, r$ e_i has a finite decomposition $e_i = e_{i1}^+ \cdots e_{ik_i}^+$, where $e_{ij}^+ \in E^+$ for $j = 1, \ldots, k_i$. Thus, $M(x, e) = \prod_{i=1}^r G(\sigma_i, e_i) = \prod_{i=1}^r G(\sigma_i, e_i^+)$. Let $x^+ = \sigma_1^+ \cdots \sigma_r^+ \in E^+$, and $e_i^+$ be any element of $G(\sigma_i, e_i^+).$ Then $p(x, e) = \pi_0 M(x, e) = \pi_0(\prod_{i=1}^r G(\sigma_i, e_i^+))$ for all $\lambda \in (0, 1), x \in T(A, \sigma, \lambda)$ iff $x^+ \in T(A, \sigma^+, \lambda)$ for all $e^+ \in Q'(e_{11}^+, \ldots, e_{1k_1}^+, e_{21}^+, \ldots, e_{r_k}^+).$ Let $e_1 = e_{i1}^+ \cdots e_{i1}^+$, $i = 1, \ldots, r$. Let $x^+ = \sigma_1^+ \cdots \sigma_r^+ \in E^*$ and let $e' \in Q'(e_{i1}^+, \ldots, e_{i1}^+, e_{i2}^+, \ldots, e_{ik}^+)$. But
again, for all $\lambda \in [0,1)$, $x \in T(A, e, \lambda)$ iff $x' \in T(A, e', \lambda)$ for all 
$e' \in Q'(e_{1,1}', \ldots, e_{k_1}', \ldots, e_{r_k}', \ldots, e_{1,1}', \ldots, e_{k_1}', \ldots, e_{r_k}')$. Thus, for all $\lambda \in [0,1)$, $x \in T(A, e, \lambda)$ 
iff $x^+ \in T(A, e^+, \lambda)$ for all $e^+ \in Q'(e_{1,1}^+, \ldots, e_{k_1}^+, \ldots, e_{r_k}^+)$ and for any 
decomposition of $e_i$ $i = 1, \ldots, r$. Note that $x^+$ depends on $x$ and the decom-
position of $e_i$ $i = 1, \ldots, r$.

Let us now consider an ADE $A^+ = (E, S, G, \pi_0^F, E^+)$. For any $x \in E^+$, 
where $L(x) = r$, we now restrict the choice of $e^+ \in Q'(e_{1,1}^+, \ldots, e_{k_1}^+, \ldots, e_{r_k}^+)$ 
$\cap (E')^*$. For any $x \in E^+$ and $e \in E^0$, such an $e^+$ must exist 
since $e_{ij}^+ \in E^+$ for $i = 1, \ldots, r$, $j = 1, \ldots, k_i$. By the above we see 
that for any $\lambda \in [0,1)$ $x \in T(A, e, \lambda)$ iff $x^+ \in T(A, e^+, \lambda)$, where $e^+$ is any 
appropriate decomposed total environment and $x^+$ is the derived input 
tape. But $\#(E') < \infty$. Hence, there is a PA $A'$ which simulates $A^+$. So 
we have a relation $R^+$ between $\Sigma^* \times E^+$ satisfying the prop-
erties of definition 2.4.1.

We now define a relation $R$ between $\Sigma^* \times E^*$ as follows: 
$[x', (x, e)] \in R$ iff for any appropriate decomposed total environment $e^+$ 
and derived input tape $x^+$, $[x', (x^+, e^+)] \in R^+$. For any $(x, e) \in E^* \times E^*$ 
we can find an associated decomposed total environment and derived input 
tape $x^+$. Hence, for any $(x, e) \in E^* \times E^*$ there exists $x' \in \Sigma^*$ such that 
$[x', (x, e)] \in R$. Also, we have for all $\lambda \in [0,1)$ that $x \in T(A, e, \lambda)$ iff 
x$\in T(A^+, e^+, \lambda)$. Hence, if $[x', (x, e)] \in R$, then for any $\lambda \in [0,1)$ 
x$\in T(A', e, \lambda)$ iff $x \in T(A, e, \lambda)$. And, clearly, $x'$ is not admissible by 
$R^+$ iff $x$ is not admissible by $R$. Thus, we have the desired result 
that $A'$ simulates $A$.

Example 2.6.1. As an example of theorem 2.6.1 let us consider 
the ADE $A_h = (\Sigma, S, G, \pi_0^F, E_h)$, where $\pi_0^F$ is a constant and
\( E_\epsilon = \{h,2h,\ldots\} \) for some constant \( h > 0 \). Clearly, \( E_\epsilon \) and the operation of addition form a semi-group. Let \( G \) be such that for any 
\[ \sigma, e \in E_\epsilon \]
\[ G(\sigma, e) = (G(\sigma, h))^{e/h} \]
Thus for any \( \sigma \in \Sigma \) and 
\[ e_1, e_2 \in E_\epsilon \]
\[ G(\sigma, e_1)G(\sigma, e_2) = (G(\sigma, h))^{e_1/h \cdot e_2/h} = (G(\sigma, h))^{(e_1+e_2)/h} \]
\[ = G(\sigma, e_1+e_2) \]
Hence, \( G \) is a homomorphism for each \( \sigma \in \Sigma \). Clearly, 
\[ E_\epsilon = S(\{h\}) \] for any \( e \in \mathcal{E}_\epsilon^* \), \( e = h+\ldots+h \); there are \( e/h \) occurrences of \( h \) in the decomposition of \( e \). Let \( x \in \mathcal{L}^* \) such that \( L(x) = r \) and 
\[ e \in \mathcal{E}_\epsilon^\infty \]
Consequently, 
\[ x^+ = \sigma_1^+ \ldots \sigma_r^+ \] and 
\[ e^+ = (h,h,\ldots) \]
Thus,
\[ M(x,e) = \prod_{i=1}^{r} G(\sigma_i, e_i) = \prod_{i=1}^{r} (G(\sigma_i, h))^{e_i/h} = \prod_{i=1}^{r} \prod_{j=1}^{r} G(\sigma_i, h) = M(x^+, e^+) \]
Now define a PA \( A' = (S', M', \pi_0', F') \) over \( \mathcal{E}' \) as in theorem 2.4.1 to simulate \( A'_\epsilon = (\mathcal{E}, S, G, \pi_0, F, \{h\}) \). Hence, \( p(x, e) = \pi_0' M'(x, e) \eta'^F = r \frac{e_i}{h} \frac{e_i}{h} \frac{e_j}{h} \frac{e_j}{h} \frac{e_k}{h} \frac{e_k}{h} \frac{e_l}{h} \frac{e_l}{h} \frac{e_m}{h} \frac{e_m}{h} \frac{\prod_{i=1}^{r} G(\sigma_i, h)}{\prod_{i=1}^{r} G(\sigma_i, h)} \eta'^F = \pi_0' M'(\phi \tau_1 \ldots \tau_r) \eta'^F = p'(\phi \tau_1 \ldots \tau_r) \]
where \( \tau_1, \ldots, \tau_r \in \mathcal{E}' \) and \( g(\sigma_i, h) = \tau_i \). Consequently, \( x \in T(A', e, \lambda) \) iff \( x' = \phi \tau_1 \ldots \tau_r \in T(A', \lambda) \) for all \( \lambda \in [0,1) \).

By our definition we may regard probabilistic automata as only being able to change state within a discrete time scale \( t = 1,2,\ldots \).
The concept of the continuous time probabilistic automata as introduced by Knast [8] is similar to the Markov chain with a continuous parameter.
The next state function of such an automaton is defined for \( t \in [0,\infty) \) as well as \( \sigma \in \Sigma \). A continuous time probabilistic automaton may change state at any time.

**Definition 2.6.2.** A continuous time probabilistic automaton (CTPA) is a system \( A = (\mathcal{E}, S, \pi_0, F, T) \), where \( \mathcal{E}, S, \pi_0 \), and \( F \) are as in definition 1.3.2, \( T = [0,\infty) \), and \( G(\sigma, t) \) is a matrix transition
function satisfying:

1. \( G(\sigma, t) \in \mathbb{M}_n \) for all \( \sigma \in \mathbb{E} \) and \( t \in T \).

2. \( G(\sigma, t_1 + t_2) = G(\sigma, t_1)G(\sigma, t_2) \) for any \( \sigma \in \mathbb{E} \) and \( t_1, t_2 \in T \).

3. Every \( g_{ij} \) is a measurable function in \((0, \infty)\) and
   \[
   \lim_{t \to 0} g_{ii}(\sigma, t) = 1 \text{ for all } \sigma \in \mathbb{E} \text{ and } i = 1, \ldots, n \text{ and } \]
   \[
   j = 1, \ldots, n.
   \]

The element \( g_{ij}(\sigma, t) \) is the probability of being in state \( s_j \) at time \( t_0 + t \) when the input is \( \sigma \) given that the system is in state \( s_i \) at time \( t_0 \) for any \( t_0 \in T \). Clearly, the transition properties depend on the time interval for input \( \sigma \).

**Definition 2.6.3.** An infinitesimal matrix of the semi-group \(((G(\sigma, t) \mid t \in T), \text{multiplication})\) is the limit \( A(\sigma) = \lim_{t \to 0} (G(\sigma, t) - I_n) / t \), where \( I_n \) denotes the \( n \times n \) identity matrix. The elements \( a_{ij}(\sigma) \) of \( A(\sigma) \) are called the rates of transition from state \( s_i \) to state \( s_j \) via input symbol \( \sigma \).

Knast proved that the transition function for every CTPA has the canonical form \( G(\sigma, t) = e^{A(\sigma)t} = \sum_{i=0}^{\infty} \frac{(tA(\sigma))^i}{i!} \).

It is readily seen from the definition of the CTPA that it is a special case of the ADE model with a semi-group environment set whose structure is preserved by the homomorphisms \( G_\sigma \) for all \( \sigma \in \mathbb{E} \). Since there does not exist a finite set \( E^+ \) such that \( T = S(E^+) \), theorem 2.5.1 may not be used to show the existence of a PA \( A' \) to simulate a CTPA. Knast showed, however, that under certain conditions a CTPA can be approximated by an ADE of the form \( A_n \) as above. For further results concerning this special case of the ADE model see Knast [8].
CHAPTER III

Stability Results

3.1 The Stability Problem

The stability problem is a natural consequence of the inevitable impreciseness of the measurement of the environment configuration. We shall consider the effects of small perturbations of the environmental condition on the acceptance sets and transition functions.

Implicit in the notion of perturbation of the environment is consideration of a natural metric on the environment set by which the size of the perturbation is measured. We shall denote the natural metric on the environment set by $d^N$. For example, let us consider a system whose transitions are related to temperature. That is, suppose $E = [0,100]; d^N(e,e') = |e-e'|$, where $e,e' \in E$, may be a useful natural metric to consider perturbations of the environment. The salient point concerning $d^N$ is that it is computed directly from the environments without necessarily considering the transition functions.

In Section 2.5 we considered the function

$$d(e,e') = \max(|\pi_0(e) - \pi_0(e')|, \max_{\sigma \in \Sigma} |G(\sigma,e) - G(\sigma,e')|)$$

based on the ADE $A = (\Sigma,\pi_0,F,E)$, where $e,e' \in E$. The function $d$ was, in general, found to be a pseudo-metric on the environment set $E$. We shall now establish the notion of continuity of the transition functions with respect to the natural metric.
Definition 3.1.1. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE and $d^N$ be a natural metric on the environment set $E$. The basic matrix transition function $G$ is continuous at $e$ with respect to $d^N$ if for every $\varepsilon > 0$ there exists $\delta(e, \varepsilon) > 0$ such that

$$\max_{\sigma \in \Sigma} |G(\sigma, e) - G(\sigma, e')| < \varepsilon$$

whenever $d^N(e, e') < \delta(e, \varepsilon)$. $G$ is uniformly continuous with respect to $d^N$ if $G$ is continuous at each $e \in E$ and $\delta$ can be chosen independent of $e$. Similarly, $\pi_0$ is continuous at $e$ with respect to $d^N$ if for every $\varepsilon > 0$ there exists $\delta(e, \varepsilon) > 0$ such that $|\pi_0(e) - \pi_0(e')| < \varepsilon$ whenever $d^N(e, e') < \delta(e, \varepsilon)$. $\pi_0$ is uniformly continuous with respect to $d^N$ if $\pi_0$ is continuous at each $e \in E$ and $\delta$ can be chosen independent of $e$.

Continuity conditions will allow us to bound the perturbations of the transition probabilities given the perturbation of the environment with respect to the natural metric.

Lemma 3.1.1. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE with natural metric $d^N$. If $G$ and $\pi_0$ are continuous at $e$, given any $\varepsilon > 0$, we can find $\delta(e, \varepsilon) > 0$ such that for any $e' \in E$, $d(e, e') < \varepsilon$ whenever $d^N(e, e') < \delta$.

Proof: Clear from the definition of continuity and $d$.

Thus, if $G$ and $\pi_0$ are continuous at $e$, for small perturbations of $e$ with respect to $d^N$ we can bound the perturbation of $G$ and $\pi_0$. Furthermore, by making the perturbation of $e$ sufficiently small we can make the perturbation of $G$ and $\pi_0$ as small as desired. Also, if $G$ and $\pi_0$ are uniformly continuous, then the choice of $\delta$ can be made independently of $e$. 

For notational convenience, let \( J = \{0,1,2,\ldots\} \) and \( J_+ = \{1,2,\ldots\} \).

**Definition 3.1.2.** An ADE \( A = (\Sigma, S, G, \pi_0, F, E) \) with natural metric \( d^N \) on \( E \) and cut-point \( \lambda \) is tape-acceptance stable (a-stable) at \( \e = (e_0,e_1,\ldots) \) iff there exists \( \delta > 0 \) such that \( \sup_{i \in J} d^N(e_i,e'_i) < \delta \) implies \( T(A,\e,\lambda) = T(A,\e',\lambda) \).

In other words, an ADE with natural metric \( d^N \) and with cut-point \( \lambda \) is a-stable if the set of tapes defined by \( A \) under total environment \( \e \) with cut-point \( \lambda \) is not changed by sufficiently small perturbations of the environments.

**Definition 3.1.3.** An ADE \( A = (\Sigma, S, G, \pi_0, F, E) \) with natural metric \( d^N \) on \( E \) is strictly stable (s-stable) at \( \e = (e_0,e_1,\ldots) \) iff given any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( \sup_{i \in J} d^N(e_i,e'_i) < \delta \) implies \( |\pi(x,\e) - \pi(x,\e')| < \varepsilon \) \( \forall x \in \mathcal{L}^* \), where \( \pi \) is the state distribution function.

**Theorem 3.1.1.** Let \( A = (\Sigma, S, G, \pi_0, F, E) \) be an ADE with natural metric \( d^N \) on \( E \). If \( A \) is s-stable at \( \e \), then given any \( \gamma > 0 \) there exists \( \delta > 0 \) such that \( \sup_{i \in J} d^N(e_i,e'_i) < \delta \) implies \( |p(x,\e) - p(x,\e')| < \gamma \) \( \forall x \in \mathcal{L}^* \), where \( p \) is the acceptance probability function.

**Proof:** \( p(x,\e) = \pi(x,\e)\eta^F \), where \( \eta^F \) is as in definition 1.3.4. For any \( x \in \mathcal{L}^* \), \( |p(x,\e) - p(x,\e')| = |(\pi(x,\e) - \pi(x,\e'))\eta^F| \leq |\pi(x,\e) - \pi(x,\e')|^{\beta(F)} \). Given any \( \gamma > 0 \) let \( \varepsilon = \gamma^{\beta(F)} \). Since \( A \) is s-stable at \( \e \), there exists \( \delta > 0 \) such that \( \sup_{i \in J} d^N(e_i,e'_i) < \delta \).
implies $|π(x,e) - π(x,e')| < ε = γ/\#(F)$. Thus, $|p(x,e) - p(x,e')| < ε/\#(F) = γ \forall x ∈ Σ^*$. 

Thus, $s$-stability at $e$ guarantees a bound on the perturbation of the acceptance probability function.

**Definition 3.1.4.** A cut-point $λ$ is called isolated with respect to an ADE $A$ at $e$ if there exists $γ > 0$ such that $|p(x,e) - λ| ≥ γ \forall x ∈ Σ^*$. 

**Theorem 3.1.2.** Let $A = (Σ,S,G,π_0,F,E)$ be an ADE with natural metric $d_N$ on $E$. If $λ$ is an isolated cut-point for $A$ at $e$ and if $A$ is $s$-stable at $e$, then $A$ with cut-point $λ$ is $a$-stable at $e$.

**Proof:** $λ$ is an isolated cut-point for $A$ at $e$. So there exists $γ > 0$ such that $|p(x,e) - λ| ≥ γ \forall x ∈ Σ^*$. Since $A$ is $s$-stable at $e$, choose $ε$ such that $0 < ε < γ$. Then by theorem 3.1.1 there exists $δ > 0$ such that $\sup_{i ∈ J} d_N(e_i, e'_i) < δ$ implies $|p(x,e) - p(x,e')| < ε < γ \forall x ∈ Σ^*$.

If $x ∈ T(A,e,λ)$, then $p(x,e) > λ$. But $λ$ is isolated, so $p(x,e) ≥ λ + γ$. Thus, $p(x,e') ≥ λ + γ - ε > λ$. Hence $x ∈ T(A,e',λ)$.

If $x ∉ T(A,e,λ)$, then $p(x,e) ≤ λ$. Again, $λ$ is isolated, so $p(x,e) ≤ λ - γ$. Thus, $p(x,e') ≤ λ - γ + ε ≤ λ$. And, hence, $x ∉ T(A,e,λ)$.

Consequently, there exists $δ > 0$ such that $\sup_{i ∈ J} d_N(e_i, e'_i) < δ$ implies $T(A,e,λ) = (A,e',λ)$. That is, $A$ with cut-point $λ$ is $a$-stable.

The concept of $a$-stability differs from $s$-stability in that $a$-stability allows a large perturbation of the acceptance probability function as long as the cut-point is not crossed.
3.2 Stability with Tapes of Bounded Length

**Definition 3.2.1.** If $\xi = (\xi_i)$ is an arbitrary vector and $P$ is an arbitrary matrix, then $||\xi|| = \sum_i |\xi_i|$ and $||P|| = \max_{i,j} |p_{ij}|$.

**Lemma 3.2.1.** a) Let $B, B', C, C' \in M_n$, then $||BC - B'C'|| \leq ||B - B'|| + ||C - C'||$. b) Let $v, v' \in V_n$, then $||vB - v'B'|| \leq ||v - v'|| + ||B - B'||$.

**Proof:** We must first show that $||\cdot||$ satisfies the triangle property on the set of all $n \times n$ matrices.

$$
||H - H'|| = \max_{i,j} |h_{ij} - h'_{ij}|
= \max_{i,j} |h_{ij} - h''_{ij} + h''_{ij} - h'_{ij}|
\leq \max_{i,j} \{|h_{ij} - h''_{ij}| + |h''_{ij} - h'_{ij}|\}
\leq \max_{i,j} \sum_{i,j} |h_{ij} - h''_{ij}| + \max_{i,j} \sum_{i,j} |h''_{ij} - h'_{ij}|
= ||H - H''|| + ||H'' - H'||$ for any $n \times n$ matrices $H, H'$, and $H''$. Let us now consider $||BC - B'C'|| = ||BC - B'C + B'C - B'C'|| \leq ||(B - B')C|| + ||B'(C - C')||$. But

$$
||(B - B')C|| = \max_{i,j,k} |\sum_{i,k} (b_{ik} - b'_{ik})c_{kj}|
\leq \max_{i,j,k} \sum_{i,k} |b_{ik} - b'_{ik}|c_{kj}
= \max_{i,j,k} \sum_{i,k} c_{kj} |b_{ik} - b'_{ik}|
= \max_{i,k} \sum_{j} |b_{ik} - b'_{ik}|
= ||B - B'||.
$$
Similarly,
\[
| |B'(C-C')| | = \max_{i,j,k} \sum \sum b'_{ik} |c_{kj} - c'_{kj}| \\
\leq \max_{i,j,k} \sum \sum b'_{ik} |c_{kj} - c'_kj| \\
= \max_{i,k} \sum b'_{ik} \sum_{j} |c_{kj} - c'_kj| \\
\leq \max_{i,k} \sum b'_{ik} | |C-C'| | \\
= | |C-C'| |.
\]
Thus, \( | |BC-B'C'| | \leq | |B-B'| | + | |C-C'| |. \)

For \( b) \ | |\cdot| | \) similarly satisfies the triangle property on the set of all \( n \)-dimensional vectors. Thus,
\[
| |vB-v'B'| | = | |vB-v'B+v'B-v'B'| |
\leq | |(v-v')B| | + | |v'(B-B')| |.
\]

We see that
\[
| |(v-v')B| | = \sum \sum_{j,i} (v_i - v'_i) b_{ij} \\
\leq \sum \sum_{j,i} b_{ij} |v_i - v'_i| \\
= \sum \sum_{i,j} b_{ij} |v_i - v'_i| \\
= \sum |v_i - v'_i| \\
= | |v-v'| |.
\]
Also,
\[
| |v'(B-B')| | = \sum \sum_{i,j} v'_i (b_{ij} - b'_{ij}) \\
\leq \sum \sum_{i,j} v'_i |b_{ij} - b'_{ij}|.
\[
= \sum_i v_i \sum_j |b_{ij} - b'_{ij}|
\leq \sum_i v_i |B-B'|
= |B-B'|.
\]

Hence, \( |vB-v'B'| \leq |v-v'| + |B-B'|. \)

\[\square\]

**Lemma 3.2.2.** For any integer \( k > 0 \), let \( B_1, \ldots, B_k \) and \( C_1, \ldots, C_k \) all belong to \( M_n \) and let \( v, w \in V_n \), then
\[
|v \prod_{i=1}^k B_i - w \prod_{i=1}^k C_i| \leq |v-w| + \sum_{i=1}^k |B_i - C_i|.
\]

**Proof:** By induction on the results of lemma 3.2.1.

**Lemma 3.2.3.** Let \( B, C \in M_n \) and \( v, w \in V_n \), then \( |B-C| \leq |B-C| \leq n|B-C| \) and \( |v-w| \leq |v-w| \leq n|v-w| \).

**Proof:** \( |B-C| = \max_{i,j} |b_{ij} - c_{ij}|. \) Let \( i^* \) and \( j^* \) be the indices \( i, j \) where the maximum is obtained. Thus, \( |B-C| = |b_{i^*j^*} - c_{i^*j^*}|. \) But
\[
|b_{i^*j^*} - c_{i^*j^*}| \leq \sum_j |b_{i^*j} - c_{i^*j}| \leq \max_{j} \sum_i |b_{ij} - c_{ij}| = |B-C|. \] And also,
\[\max_{i,j} \sum |b_{ij} - c_{ij}| \leq \max_{i,j} \sum |b_{i^*j^*} - c_{i^*j^*}| = n|B-C|. \] And similarly for the result concerning vectors.

\[\square\]

Keeping in mind the above results, let us now consider an \( ADE \)
\[A = (\Sigma, S, G, \pi_0, F, E)\] with natural metric \( d^N \) on \( E. \) For any fixed \( K, \)
let \( x \in E^* \) be an arbitrary input tape such that \( L(x) = k \leq K. \) If \( x \) is input under some environment \( e = (e_0, e_1, \ldots) \), we evaluate
\[p(x, e) = \pi_0(e_0)(\prod_{i=1}^k G(a, e_i))n^F. \] Let us now consider the acceptance probability when \( e_0, \ldots, e_K \) are perturbed with respect to \( d^N \) by less than \( \delta. \) That is, \( \max_{i=0, \ldots, K} d^N(e_i, e_i') < \delta. \) If we assume that \( \pi_0 \) is
continuous at $e_0$ and $G$ is continuous at each $e_i$, $i = 1, \ldots, K$, then

$$|\pi_0(e_0) - \pi_0(e'_0)| < \varepsilon_1(e_0)$$

and

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \varepsilon_2(e_i),$$

for $i = 1, \ldots, K$. Thus, if we let

$$\varepsilon^* = \max_i (\varepsilon_1(e_0), \max_{i=1, \ldots, K} \varepsilon_2(e_i)),$$

then

$$\max_{i=0, \ldots, K} d^N(e_i, e'_i) < \delta$$

implies

$$|\pi_0(e_0) - \pi_0(e'_0)| < \varepsilon^*$$

and

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \varepsilon^*, \quad i = 1, \ldots, K.$$ Using lemma 3.2.3 we see that

$$\max_{i=0, \ldots, K} d^N(e_i, e'_i) < \delta$$

implies

$$||\pi_0(e_0) - \pi_0(e'_0)|| < n\varepsilon^*$$

and

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < n\varepsilon^*, \quad i = 1, \ldots, K.$$ By lemma 3.2.2 we have

$$||\pi(x, e) - (x, e')|| = ||\pi_0(e_0) - \pi_0(e'_0)|| + \sum_{i=1}^{k} ||G(\pi, e_i) - G(\pi, e'_i)||$$

$$\leq ||\pi_0(e_0) - \pi_0(e'_0)|| + \sum_{i=1}^{k} ||G(\sigma, e_i) - G(\sigma, e'_i)||$$

$$\leq n\varepsilon^* + k n \varepsilon^*$$

$$\leq (K+1)n\varepsilon^*$$

such that $L(x) \leq K$.

and $\exists e' \in E^*$ such that

$$\max_{i=0, \ldots, K} d^N(e_i, e'_i) < \delta.$$ Now, consider

$$|p(x, e) - p(x, e')| = |(\pi(x, e) - \pi(x, e'))_E|$$

$$\leq \sum_i |\pi(x, e) - \pi(x, e')_i|$$

$$= ||\pi(x, e) - \pi(x, e')||.$$

Thus,

$$\max_{i=0, \ldots, K} d^N(e_i, e'_i) < \delta$$

implies

$$||\pi(x, e) - \pi(x, e')|| \leq (K+1) n \varepsilon^*$$

and

$$|p(x, e) - p(x, e')| \leq (K+1) n \varepsilon^*$$

for any $x \in E^*$ such that $L(x) \leq K$. 
Since $\kappa$ and $n$ are fixed finite numbers, given any $\varepsilon > 0$, we let $0 < \varepsilon^* < \varepsilon/(K+1)n$. By the continuity of $\pi_0$ at $e_0$ and $G$ at each $e_i$ $i = 1,\ldots,\kappa$ there exists $\delta > 0$ such that

$$\max_{i=0,\ldots,K} d^N(e_i, e'_i) < \delta \quad \text{implies} \quad |\pi_0(e_0) - \pi_0(e'_0)| < \varepsilon^*$$

and

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \varepsilon^* \quad \text{for} \quad i = 1,\ldots,K. \quad \text{In turn, this implies}$$

$$|p(x, e) - p(x, e')| \leq (\kappa+1)n \varepsilon^* < \varepsilon \quad \forall x \in \Sigma^* \quad \text{such that} \quad L(x) \leq K.$$

**Theorem 3.2.1.** Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE with natural metric $d^N$ on $E$. For any fixed $e = (e_0, e_1,\ldots) \in E^\infty$ and finite non-negative integer $K$, if

1. $\pi_0$ is continuous with respect to $d^N$ at $e_0$;
2. $G$ is continuous with respect to $d^N$ at each $e_i$ $i = 1,\ldots,K$

then given any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\max_{i=0,\ldots,K} d^N(e_i, e'_i) < \delta \quad \text{implies} \quad |p(x, e) - p(x, e')| < \varepsilon$$

for all $x \in \Sigma^*$ such that $L(x) \leq K$.

When the continuity conditions are satisfied, we can apply theorem 3.2.1 to say that for any $K$ and any $x \in \Sigma^*$ such that $L(x) \leq K$, $p(x, e') + p(x, e)$ as $\max_{i=0,\ldots,K} d^N(e_i, e'_i) \to 0$. Thus, $p(x, e')$ can be used to approximate $p(x, e)$ and we have a bound on the maximum error of the approximation as a function of the largest perturbation of $e_0, e_1,\ldots, e_K$.

**Example 3.2.1.** Consider the ADE $A$ with $E = \{0, 1,\ldots, m-1\}$, $S = \{s_1, s_2, s_3\}$, and $E = [0, 1]$. The basic matrix transition function is
and the initial distribution function is \( \pi_0(e) = (e, 1-e, 0) \). Let \( F = \{ s_3 \} \); hence, \( \eta_F^F = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \). Also let \( e = (1, 1, \ldots) \).

Suppose Euclidean distance is the natural metric on \( E \). Thus, we have

\[ |\pi_0(e') - \pi_0(1) - (1-e')| = d^N(1, e') \quad \text{and} \quad \max_{\sigma \in \Sigma} |G(\sigma, e') - G(\sigma, 1)| \leq |1-e'| = d^N(1, e'). \]

So then \( \pi_0 \) and \( g \) are continuous at \( e = 1 \).

Let \( K \) be any nonnegative integer and \( x \in \mathbb{E}^* \) such that \( L(x) \leq K \).

Given any \( \varepsilon > 0 \), we can find \( \delta > 0 \) such that for every \( e' \in \mathbb{E}^\infty \) that satisfies

\[ \max_{i=0, \ldots, K} d^N(1, e_1^i) = \max_{i=0, \ldots, K} |1-e_1^i| < \delta \]

we have

\[ |p(x, e^i) - p(x, (1, 1, \ldots))| < \varepsilon. \]

In fact, we can see that for

\[ L(x) = k \leq K, \quad p(x, e^i) = \pi_0(e^i) \left( \sum_{i=1}^k G(\sigma_i, e^i) \right) \eta_F^F = e_0^i \sum_{i=1}^k \frac{\sigma_i e^i}{i}. \]

And similarly,

\[ p(x, (1, 1, \ldots)) = \sum_{i=1}^k \frac{\sigma_i}{i}. \]

So we choose \( \delta < \varepsilon/(K+1) \) and we have

\[ |p(x, e^i) - p(x, (1, 1, \ldots))| = \left| e_0^i \sum_{i=1}^k \frac{\sigma_i e^i}{i} - \sum_{i=1}^k \frac{\sigma_i}{i} \right| \leq \frac{k}{i=0}^k |e'_i - 1| \quad \text{and} \quad \sum_{i=0}^k |e'_i - 1| \leq \delta(K+1) < \varepsilon. \]

If we consider \( A \) operating under the environment \( (1, 1, \ldots) \), we see that such a system is a probabilistic automaton \( A^1 = (S, M, \pi_0, F) \) over \( \Sigma \), where \( S, F \), and \( \Sigma \) are in \( A \) and \( \pi_0 = \pi_0(1) \) and \( M(\sigma) = G(\sigma, 1) \). So for any \( \lambda \in [0, 1] \), \( T(A^1, \lambda) = T(A, (1, 1, \ldots), \lambda) \) since \( p^1(x) = p(x, (1, 1, \ldots)) \forall x \in \mathbb{E}^* \).

Hence, we can use the probabilistic automaton \( A^1 \) to approximate

\[
G(\sigma, e) = \begin{pmatrix}
1/m & m - \sigma e - 1/m & \sigma e/m \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]
the acceptance probability of any input for $A$. Moreover, for any nonnegative integer $K$, any $\lambda \in [0,1)$, and any $x \in \Sigma^*$, $p(x, e(\delta,K))$

\[ \leq p(x, e') \leq p^1(x), \text{ where } \max_{i=0,\ldots,K} d^N(e'_i, l) < \delta \text{ and } e(\delta,K) = (1-\delta,\ldots,1-\delta,0,0,\ldots). \]

Hence, $T(A, e(\delta,K), \lambda) < T(A, e', \lambda) < T(A^1, \lambda)$.

But $T(A, e(\delta,K), \lambda)$ is strictly increasing as $\delta \to 0$ and bounded above by $T(A^1, \lambda)$. So the limit exists.

\[ \lim_{\delta \to 0} T(A, e(\delta,K), \lambda) = \{ x \in \Sigma^* | x \in T(A^1, \lambda) \text{ and } L(x) \leq K \}. \]

Also, $\lim_{\delta \to 0} T(A, e(\delta,K), \lambda)$ is strictly increasing as $K \to \infty$ and bounded above by $T(A^1, \lambda)$; hence, this limit exists.

\[ \lim_{K \to \infty} \lim_{\delta \to 0} T(A, e(\delta,K), \lambda) = \lim_{K \to \infty} \{ x \in \Sigma^* | x \in T(A^1, \lambda) \text{ and } L(x) \leq K \} = T(A^1, \lambda) . \]

Similarly, $\lim_{\delta \to 0} \lim_{K \to \infty} T(A, e(\delta,K), \lambda) = T(A^1, \lambda)$.

Now we can say that

\[ \lim_{K \to \infty} \max_{i=0,\ldots,K} d^N(e'_i, l) \to 0 \]

for all $\lambda \in [0,1)$.

3.3 Ergodic Conditions for $s$-Stability

In the previous section we produced sufficient conditions for a stability result on sets of input tapes of bounded length. We shall now consider the asymptotic properties of long products of stochastic
matrices to generate sufficient conditions for $s$-stability. That is, under such conditions for some fixed $e \in \mathbb{R}^\infty$, given any $e > 0$, we can find $\delta > 0$ such that if the largest perturbation of the environment is less than $\delta$, then the largest perturbation of the state distribution function is less than $e$ for any input tape.

**Definition 3.3.1.** For any real number $a$, let $a^+ = \max(a, 0)$ and $a^- = \min(a, 0)$. For any $p \in \mathcal{P}_n$, we define

$$
\delta(p) = \max_{i_1, i_2} \sum_{j} (p_{i_1j} - p_{i_2j})^+.
$$

**Lemma 3.3.1.** For any $p, q \in \mathcal{P}_n$, $\sum_{i=1}^{n} (p_{i} - q_{i})^+ = -\sum_{i=1}^{n} (p_{i} - q_{i})^-$

$$
= \frac{1}{2} \sum_{i=1}^{n} |p_{i} - q_{i}|.
$$

**Proof:** Let $N' = \{i | p_i - q_i \geq 0\}$ and $N'' = \{i | p_i - q_i < 0\}$. Then $N' \cup N'' = \{1, 2, \ldots, n\}$ and $N' \cap N'' = \emptyset$

$$
0 = \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} (p_i - q_i) = \sum_{i \in N'} (p_i - q_i) + \sum_{i \in N''} (p_i - q_i)
$$

$$
= \sum_{i=1}^{n} (p_i - q_i)^+ + \sum_{i=1}^{n} (p_i - q_i)^-.
$$

Hence, $\sum_{i=1}^{n} (p_i - q_i)^+ = -\sum_{i=1}^{n} (p_i - q_i)^-$. Also,

$$
\frac{1}{2} \sum_{i=1}^{n} |p_{i} - q_{i}| = \frac{1}{2} \sum_{i=1}^{n} (p_{i} - q_{i})^+ - \frac{1}{2} \sum_{i=1}^{n} (p_{i} - q_{i})^- = \frac{1}{2} \sum_{i=1}^{n} (p_{i} - q_{i})^+ + \frac{1}{2} \sum_{i=1}^{n} (p_{i} - q_{i})^+
$$

$$
= \sum_{i=1}^{n} (p_{i} - q_{i})^+ = -\sum_{i=1}^{n} (p_{i} - q_{i})^-.
$$

$\square$
Lemma 3.3.2. For any $A, B \in M_n$,

1. $0 \leq \varepsilon(A) \leq 1$

2. $\varepsilon(AB) \leq \varepsilon(A)\varepsilon(B)$.

Proof: Let us consider any fixed indices $i_1$ and $i_2$, then

$$0 \leq \sum_j (a_{i_1j} - a_{i_2j})^+ \leq \sum_j a_{i_1j} = 1.$$ Hence $0 \leq \varepsilon(A) = \max_{i_1, i_2} \sum_j (a_{i_1j} - a_{i_2j})^+ \leq 1$.

Again, let us consider any fixed indices $i_1$ and $i_2$. Let

$N' = \{j | \sum_k (a_{i_1k} - a_{i_2k})b_{kj} \geq 0\}$. Now we have

$$\sum_j (\sum_k (a_{i_1k} - a_{i_2k})b_{kj})^+ = \sum_j \sum_k (a_{i_1k} - a_{i_2k})b_{kj}$$

$$= \sum_k (a_{i_1k} - a_{i_2k}) \sum_j b_{kj}$$

$$\leq \sum_k (a_{i_1k} - a_{i_2k})^+ \max_k \sum_j b_{kj}$$

$$+ \sum_k (a_{i_1k} - a_{i_2k})^- \min_k \sum_j b_{kj}.$$ By lemma 3.3.1

$$\sum_j (\sum_k (a_{i_1k} - a_{i_2k})b_{kj})^+ \leq \sum_k (a_{i_1k} - a_{i_2k})^+ \max_k \sum_j b_{kj} - \min_k \sum_j b_{kj}$$

Note that the elements of $N'$ depend on $i_1$ and $i_2$, but

$$\max_k \sum_{j \in N'} b_{kj} - \min_k \sum_{j \in N'} b_{kj} = \max_{k_1, k_2} \sum_{j \in N'} (b_{k_1j} - b_{k_2j})$$

$$\leq \max_{k_1, k_2} \sum_{j \in N'} (b_{k_1j} - b_{k_2j})^+$$

$$\leq \varepsilon(B) = \varepsilon(B)$$
which is independent of $i_1$ and $i_2$. Therefore,

$$
\sum_j \left( \sum_k (a_{1k} - a_{2k}) b_{kj} \right)^+ \leq \sum_k (a_{1k} - a_{2k})^+ \alpha(B).
$$

Now,

$$
\alpha(AB) = \max_{i_1, i_2} \left( \sum_k (a_{i_1k} - a_{i_2k}) b_{kj} \right)^+ \\
\leq \max_{i_1, i_2} \sum_k (a_{i_1k} - a_{i_2k})^+ \alpha(B) \\
= \alpha(A) \alpha(B).
$$

\[\square\]

**Lemma 3.3.3.** For any $B \in M_n$ and any $n$-dimensional vector $\xi$ such that $\sum_i \xi_i = 0$, then $||\xi B|| \leq ||\xi|| \alpha(B)$.

**Proof:** If $\xi$ is the zero vector, then $0 = ||\xi B|| = ||\xi|| \alpha(B)$ and the result is true.

Let $\xi$ be any nonzero vector such that $\sum_i \xi_i = 0$. Define the vectors $\xi^1$ and $\xi^2$ as

$$
\xi_1^1 = \frac{2\xi}{||\xi||} \text{ and } \xi_1^2 = -\frac{2\xi}{||\xi||}.
$$

Since $\sum_i \xi_i = 0$, we have that $\sum_i \xi_i^+ = -\sum_i \xi_i^- = 1/2 \sum_i |\xi_i|$. Hence

$$
\xi_1^1, \xi_2^2 \in V_n \text{ and } \xi_1^1 - \xi_2^2 = \frac{2\xi}{||\xi||}.
$$

Let $H \in M_n$ be such that its first row is $\xi_1^1$ and all of its other rows are $\xi_2^2$. So $2\alpha(HB) = 2 \sum_j \left( \sum_k (\xi_1^1_k - \xi_2^2_k) b_{kj} \right)^+$. By lemma 3.3.1,
\[ 2 \vartheta(HB) = \sum_{j} \sum_{k} (\xi_{k}^{1} - \xi_{k}^{2}) b_{kj} \leq 2 \sum_{j} \sum_{k} \frac{\xi_{k} b_{kj}}{||\xi||} = \frac{2}{||\xi||} \sum_{k} ||\xi_{k} b_{kj}|| = 2 \frac{||\xi B||}{||\xi||}. \]

Thus, \( \frac{||\xi B||}{||\xi||} = \vartheta(HB) \leq \vartheta(H) \vartheta(B) \leq \vartheta(B) \) by lemma 3.3.2. And therefore, the result \( ||\xi B|| \leq ||\xi|| \vartheta(B). \) \]

**Lemma 3.3.4.** For any \( B \in M_{n} \) and any \( n \times n \) matrix \( A \) such that all of its rows satisfy the condition for \( \xi \) in lemma 3.3.3, then \( ||AB|| \leq ||A|| \vartheta(B). \)

**Proof:** Let \( \alpha_{i}^{1} \) denote the \( i \)-th row of \( A. \)

\[
||AB|| = \max_{i} \sum_{k} a_{ik}^{1} b_{jk} \leq \vartheta(B) \max_{i} \sum_{k} |a_{ik}^{1}| = \vartheta(B) \max_{i} |a_{i}^{1}| = \vartheta(B) \max_{i} \sum_{j} |a_{ij}| = ||A|| \vartheta(B). \]

**Lemma 3.3.5.** Let \( A, B \in M_{n} \), then \( ||AB-B|| \leq 2 \vartheta(B) . \)

**Proof:** \( ||AB-B|| = ||(A-I_{n})B||. \) Each row of \( A-I_{n} \) sums to 0 and \( ||A-I_{n}|| \leq 2. \) Hence, by lemma 3.3.4 \( ||AB-B|| \leq 2 \vartheta(B). \)

**Definition 3.3.2.** An ADE \( A = (E,S,G,\pi_{0},F,\xi) \) is weakly ergodic (\( \omega \)-ergodic) iff for any \( \epsilon > 0 \) there exists \( N(\epsilon) \) such that \( \vartheta(M(x,\epsilon)) < \epsilon \) \( \forall \epsilon \in E^{\infty} \) and \( \forall x \in E^{*} \) such that \( L(x) \geq N(\epsilon). \)

The weakly ergodic condition is a generalization of quasideterminacy for probabilistic automata. That is, for any two stochastic vectors
\(\pi\) and \(\rho\), we have 
\[|(\pi - \rho)M(x, e)| \leq |(\pi - \rho)M(x, e)| \leq 2\varepsilon(M(x, e)) \leq 2\varepsilon\] 
for all \(x\) such that \(L(x) \geq N(\varepsilon)\). If we have an ADE that is weakly ergodic, then \(\varepsilon(M(x, e)) \to 0\) as \(L(x) \to \infty\) uniformly with respect to \(e\).

**Definition 3.3.3.** Let \(e = (e_0, e_1, \ldots) \in E^\infty\). The total environment \(e_k = (e_k, e_{k+1}, \ldots) \in E^\infty\) is called the \(k\)-tail of \(e\). \(E^k(e) = \{e_k | k \in J\}\) is called the set of all tails of \(e\).

**Theorem 3.3.1.** An ADE \(A\) is \(w\)-ergodic iff there exists \(k \in J\) and \(0 < \delta < 1\) such that 
\[\varepsilon(M(x, e)) < \delta \quad \forall e \in E^\infty \quad \text{and} \quad \forall x \in E^k \quad \text{such that} \quad L(x) = k.\]

**Proof:** If \(A\) is \(w\)-ergodic, then for any \(\varepsilon > 0\), there exists \(N(\varepsilon)\) such that \(\varepsilon(M(x, e)) < \varepsilon \quad \forall e \in E^\infty \quad \text{and} \quad \forall x \in E^k \quad \text{such that} \quad L(x) = k.\) Choose any \(0 < \varepsilon < 1\), then let \(\delta = \varepsilon\) and \(k = N(\varepsilon)\).

For the converse, suppose there exist \(k \in J\) and \(0 < \delta < 1\) such that 
\[\varepsilon(M(x, e)) < \delta \quad \forall e \in E^\infty \quad \text{and} \quad \forall x \in E^k \quad \text{such that} \quad L(x) = k.\] For any \(\varepsilon > 0\) we can find \(n_0\) such that \(\delta^{n_0} < \varepsilon\). Let \(x\) be any tape such that \(L(x) = kn_0\). Then \(x\) has the form \(x = y_1, \ldots, y_n, y_{n_0}\), where \(L(y_1) = k\) and \(L(y) = 0\). Thus, for any fixed, but arbitrary \(e \in E^\infty\), \(M(x, e) = (\prod_{i=1}^{n_0} M(y_i, e_{k(i-1)}))M(y, e_{n_0,k})\). And by lemma 3.3.2, 
\[\varepsilon(M(x, e)) \leq (\prod_{i=1}^{n_0} \varepsilon(M(y_i, e_{k(i-1)})))\varepsilon(M(y, e_{n_0,k})) \leq \delta^{n_0}\varepsilon(M(y, e_{n_0,k})).\]

Also by lemma 3.3.2, \(0 \leq \varepsilon(M(y, e_{n_0,k})) \leq 1\), so \(\varepsilon(M(x, e)) \leq \delta^{n_0} < \varepsilon\).

Thus, given any \(\varepsilon > 0\), choose \(N(\varepsilon) = kn_0\). \(\square\)
Suppose we have an ADE $A$ such that for each $e \in E^\infty$, given any $\epsilon > 0$, we can find $N(\epsilon, e) > 0$ such that $\vartheta(M(x, e)) < \epsilon \forall x \in E^\star$ such that $L(x) \geq N(\epsilon, e)$. We present the following example to demonstrate that such an ADE need not be w-ergodic.

**Example 3.3.1.** Let $A$ with $E = \{\sigma\}$, $E = (0, 1)$, and

$$G(\sigma, e) = \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix}.$$ The matrix function $G(\sigma, e)$ can be written as

$$G(\sigma, e) = U_1 + (1-2e)U_2,$$ where

$$U_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$ and

$$U_2 = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$ But $U_1^2 = U_1$, $U_2^2 = U_2$, $U_1U_2 = U_2U_1 = 0$. So by induction we have that

$$\prod_{i=1}^k G(\sigma, e_i) = U_1 + U_2\left(\prod_{i=1}^k (1-2e_i)\right).$$

Thus

$$\vartheta\left(\prod_{i=1}^k G(\sigma, e_i)\right) = (U_1 + U_2\left(\prod_{i=1}^k (1-2e_i)\right)) = \prod_{i=1}^k |1-2e_i|.$$ Now, for any $e = (e_0, e_1, \ldots) \in E^\infty$ and any $x \in E^\star$ such that $L(x) = k$, $H(x, e) = \prod_{i=1}^k G(\sigma, e_i)$. Hence, $\vartheta(H(x, e)) = \prod_{i=1}^k |1-2e_i|$, where $L(x) = k$. But $|1-2e| < 1 \forall e \in E$. So given any $e \in E^\infty$ and any $\epsilon > 0$, we can find $N(\epsilon)$ such that $\prod_{i=1}^N |1-2e_i| < \epsilon$. Thus, for this $e \in E^\infty$, $\vartheta(H(x, e)) = \prod_{i=1}^k |1-2e_i| \leq \prod_{i=1}^N |1-2e_i| < \epsilon \forall x \in E^\star$ such that $L(x) = k \geq N(\epsilon)$. However, this choice of $N(\epsilon)$ depends on $e$. And for any $N(\epsilon)$, we can find $e' \in E^\infty$ such that $\prod_{i=1}^N |1-2e'_i|$ is arbitrarily close to 1. So for some $0 < \epsilon < 1$ and any $N(\epsilon)$, there is $e' \in E^\infty$ and $x \in E^\star$ such that $L(x) \geq N(\epsilon)$ whereby $\vartheta(H(x, e')) > \epsilon$. Hence, $A$ is not w-ergodic.
Definition 3.3.4. An ADE $A = (E, S, G, \pi_0 F, E)$ is ever-ergodic at $e \in E^\infty$ iff for any $\epsilon > 0$ there exists $N(\epsilon)$ such that $\delta(M(x, e_k)) < \epsilon$ $\forall e_k \in E^\epsilon(\epsilon)$ and $\forall x \in E^\epsilon$ such that $L(x) \geq N(\epsilon)$.

Theorem 3.3.2. If $A$ is an ADE that is w-ergodic, then for each $e \in E^\infty$, $A$ is ever-ergodic at $e$.

Proof: If $A$ is w-ergodic, then given any $\epsilon > 0$, there exists $N(\epsilon)$ such that $\delta(M(x, e)) < \epsilon$ $\forall e \in E^\infty$ and $\forall x \in E^\epsilon$ such that $L(x) \geq N(\epsilon)$. But for any $e \in E^\infty$, we have $E^\epsilon(\epsilon) \subset E^\infty$. Then, clearly, given any $\epsilon > 0$, choose $N(\epsilon)$ as above; and hence, $\delta(M(x, e_k)) < \epsilon$ $\forall e_k \in E^\epsilon(\epsilon)$ and $\forall x \in E^\epsilon$ such that $L(x) \geq N(\epsilon)$. $\square$

The w-ergodic property is global in $E^\infty$ while the ever-ergodic property is a pointwise property. If the choice of $N(\epsilon)$ is not dependent on $e$ and if $A$ is ever-ergodic at each point of $E^\infty$, then $A$ is w-ergodic.

Lemma 3.3.6. If $A$ is an ADE that is ever-ergodic at $e$, then $A$ is ever-ergodic at $e_k$ $\forall e_k \in E^\epsilon(\epsilon)$. Further, if there exists $k \in J$ such that $A$ is ever-ergodic at $e_k$, then $A$ is ever-ergodic at $e_k$.

Proof: Let $A$ be ever-ergodic at $e$. For all $k \in J$, $E^\epsilon(e_k) \subset E^\epsilon(e)$. Clearly, this implies $A$ is ever-ergodic at $e_k$ $\forall k \in J$. Also, given any $\epsilon > 0$, the choice $N(\epsilon)$ need not depend on $k$.

Suppose there exists $k \in J$ such that $A$ is ever-ergodic at $e_k$. Therefore, given any $\epsilon > 0$, there exists $N_k(\epsilon)$ such that $\delta(M(e_{k+j})) < \epsilon$ $\forall j \in J$ and $\forall x \in E^\epsilon$ such that $L(x) \geq N_k(\epsilon)$. For this $\epsilon$
to prove ever-ergodicity at $\epsilon$ we simply choose $N(\epsilon) = N_k(\epsilon) + k$.

Let $x \in \mathbb{E}^*$ such that $L(x) \geq N(\epsilon) = N_k(\epsilon) + k$. So $x = x_k y$, where $L(x_k) = k$ and $L(y) \geq N_k(\epsilon)$. Thus, $\mathbb{M}(x, \epsilon) = \mathbb{M}(x_k, \epsilon) \mathbb{M}(y, \epsilon_k)$. Now by lemma 3.3.2

$$\mathbb{E}(\mathbb{M}(x, \epsilon)) \leq \mathbb{E}(\mathbb{M}(x_k, \epsilon)) \mathbb{E}(\mathbb{M}(y, \epsilon_k))$$

$$\leq \mathbb{E}(\mathbb{M}(y, \epsilon_k)) < \epsilon.$$ 

Lemma 3.3.7. An ADE $A$ is ever-ergodic at $\epsilon$ iff there exists $0 < \delta < 1$ and an increasing sequence of nonnegative integers $(k_0, k_1, \ldots)$, where $k_0 = 0$ and $k_{i+1} - k_i \leq K$ for some $K$, such that $\mathbb{E}(\mathbb{M}(x_i, \epsilon_{k_i})) \leq \delta$

$\forall x_i \in \mathbb{E}^*$ such that $L(x_i) = k_{i+1} - k_i$.

Proof: If $A$ is ever-ergodic at $\epsilon$, then given any $0 < \epsilon < 1$, there exists $N(\epsilon)$ such that $\mathbb{E}(\mathbb{M}(x, \epsilon_{j})) < \epsilon$ $\forall x_j \in \mathbb{E}^*(\epsilon)$ and $\forall x \in \mathbb{E}^*$ such that $L(x) \geq N(\epsilon)$. Choose $\delta = \epsilon$, $K = N(\epsilon)$, and $k_1 = iK$.

Suppose now we have $0 < \delta < 1$ and a sequence $(k_0, k_1, \ldots)$ as above. Therefore now we have $0 < \delta < 1$ and a sequence $(k_0, k_1, \ldots)$ as above. Therefore given any $\epsilon > 0$, there exists $n_0$ such that

$\delta^{n_0} < \epsilon$. We wish to examine the behavior of $\mathbb{M}$ at an arbitrary tail of

$e, e_j$, where $j$ is any fixed, but arbitrary positive integer. We can find $j^* \in J$ such that $k_{j^*+j}$ and $k_{j^*+1} \leq j$. Let $x \in \mathbb{E}^*$ such that $L(x) \geq (n_0+1)k$. Thus, $x$ has the form $x = y_1 x_1 \cdots x_n y_2$, where $L(x_1) = k_{j^*+1} - k_{j^*+1-1} \leq K$,

$L(y_1) = k_{j^*+1} - k \leq K$, and $L(y_2) \geq n_0 + 1$. Hence,

$$\mathbb{M}(x, e_j) = \mathbb{M}(y_1, e_j) \left( \prod_{i=j^*}^{j^{n_0}+1} \mathbb{M}(x_{i-j^{n_0}+1}, e_{k_i}) \right) \mathbb{M}(y_2, e_{k_{j^*+n_0+1}}).$$

So by lemma 3.3.2 we deduce
\[ \exists (M(x,e_j) \leq \exists (\exists (y_{1\cdot e_j}) \left( \prod_{i=j}^{j+n_0} \exists (M(x_{1\cdot j}^{i}, e_{j}) \right) \exists (C_2, e_{j}^{j+n_0} + 1)) \leq \delta^{n_0} < \epsilon. \]

Since \( j \) is an arbitrary positive integer, given any \( \epsilon > 0 \), we can find \( N(\epsilon) = (n_0+1)K \) and \( n_0 \) such that \( \exists (M(x,e_j)) < \epsilon \quad \forall e_j, \quad j=1,2,... \)
and \( \forall x \in \Sigma^* \) such that \( L(x) \geq N(\epsilon) \). That is, \( A \) is ever-ergodic at \( e_1 \). Hence, by lemma 3.3.6 \( A \) is ever-ergodic at \( e_1 \).

**Definition 3.3.5.** For some \( \delta > 0 \), a matrix \( A \in M_n \) is \( \delta \)-scrambling iff for any pair of rows \( i_1 \) and \( i_2 \), there is a column \( j \) such that \( a_{i_1 j} > \delta \) and \( a_{i_2 j} > \delta \).

**Definition 3.3.6.** An ADE \( A = (\Sigma, S, G, \pi_0, F, E) \) is said to satisfy the ever-\( H_4(\delta) \) condition at \( e \) iff there exists \( \exists (\delta) \) such that \( \forall x \in \Sigma^* \), where \( L(x) \geq \exists (\delta) \), \( M(x,e_k) \) is \( \delta \)-scrambling \( \forall e_k \in E^*(\epsilon) \).

These definitions are extensions of the concepts concerning probabilistic automata. Note that if \( A \in M_n \) is \( \delta \)-scrambling, then \( \exists (A) < 1-\delta \). We now extend theorem 1.3.6.

**Theorem 3.3.3.** \( A \) is ever-ergodic at \( e \) iff for some \( \delta > 0 \), \( A \) satisfies the ever-\( H_4(\delta) \) condition at \( e \).

**Proof:** If \( A \) is ever-ergodic at \( e \), then given any \( \epsilon > 0 \), there exists \( N(\epsilon) \) such that \( \exists (M(x,e_k)) < \epsilon \quad \forall e_k \in E^*(\epsilon) \) and \( \forall x \in \Sigma^* \) such that \( L(x) \geq N(\epsilon) \). For some \( k \in J \), let \( i_1 \) be the index of an arbitrary row \( M(x,e_k) \). Since \( M(x,e_k) \in M_n \), then there exists
s_j \in S such that \( m_{i,j}(x, e_k) \geq 1/n \). But \( \exists (M(x, e_k)) < \epsilon \), so for any 
\( i_2 \neq i_1, \ |m_{i,j}(x, e_k) - m_{i_2,j}(x, e_k)| < \epsilon \). Hence, if we choose \( \epsilon = 1/2n \),
then \( m_{i,j}(x, e_k) > 1/2n \) for any \( i \). So, choose \( \delta = 1/2n \) and 
\( N(1/2n) = \tilde{N}(1/2n) \). The choice of \( k \) is arbitrary. Consequently,
\( M(x, e_k) \) is \( 1/2n \)-scrambling \( \forall e_k \in E^*(e) \) and \( \forall e \in E^* \) such that 
\( L(x) \geq \tilde{N}(1/2n) \).

Suppose now that we have \( \delta > 0 \) such that \( A \) satisfies the 
ever-H_4(\delta) condition at \( e \). Thus, there exists \( M(\delta) \) such that \( \forall e \in E^* \)
such that \( L(x) \geq N(\delta) \) implies \( M(x, e_k) \) is \( \delta \)-scrambling \( \forall e_k \in E^*(e) \).
Hence, \( \exists (M(x, e_k)) = 1 - \delta < 1 \) \( \forall e_k \in E^*(e) \) and \( \forall e \in E^* \) such that 
\( L(x) \geq \tilde{N}(\delta) \). Let \( k_i = i(\delta) \) and \( 0 < \delta^* = 1-\delta < 1 \). By lemma 3.3.7
\( A \) is ever-ergodic at \( e \).

We shall now see that under a certain continuity condition, if 
\( A \) is ever-ergodic at \( e \), then \( A \) is ever-ergodic at each \( e' \) in 
some neighborhood of \( e \) with respect to the pseudo-metric \( \sup_{i \in J_+} d^{N(e_i, e'_i)} \)
on \( \tilde{E} \).

**Theorem 3.3.4.** If \( A \) is ever-ergodic at \( e \in E^\infty \), then there exists 
\( \epsilon > 0 \) such that for any \( e' \in E^\infty \), \( A \) is ever-ergodic at \( e' \) if 
\( \max_{\sigma \in E} ||G(\sigma, e_i) - G(\sigma, e'_i)|| < \epsilon \) for all \( i \in J_+ \).

**Proof:** Let \( k \in J \) be fixed, but arbitrary. Let \( M^{i_0}(x, e_k) \) denote 
the matrix in \( M_i \) all of whose rows are equal to the \( i_0 \)-th row 
of \( M(x, e_k) \). Then by lemma 3.3.1 
\( ||M(x, e_k) - M^{i_0}(x, e_k)|| = \)
\( \max_{i} \sum_{j} |m_{i,j} - m_{i_0,j}| = 2 \max_{j} \sum_{i} (m_{i,j} - m_{i_0,j})^+ \), where \( M(x, e_k) = (m_{ij}) \).
Hence,
\[ \| i(x, e_k) - i_0(x, e_k) \| \leq 2 \max_{i_1, i_2, j} \sum_{m_{i_1} j - m_{i_2} j}^+ = 2 \hat{a}(M(x, e_k))\]

Since \( A \) is ever-ergodic at \( e \), for any \( 0 < \delta < 1 \) there exists \( n_0 \) independent of \( k \) such that \( \hat{a}(i(x, e_k)) < \delta/6 \), \( \forall e_k \epsilon E^*(e) \) and \( \forall x \epsilon E^* \) satisfying \( L(x) \geq n_0 \). Thus, by the above proved inequality, we have
\[ \| i_0(x, e_k) - i_0(x, e_k) \| < \delta/3 \forall s_{i_0} \epsilon S \text{ and } \forall x \epsilon E^* \text{ such that } L(x) \geq n_0. \]

Let \( \hat{x} \) be a fixed, but arbitrary tape such that \( L(\hat{x}) = n_0. \)

Hence,
\[ \| i_0(\hat{x}, e_k) - i_0(\hat{x}, e_k) \| < \delta/3 \forall s_{i_0} \epsilon S. \]

Now let \( 0 < \epsilon < \delta/3n_0 \). Hence, for any \( s_{i_0} \epsilon S \),
\[ \| i_0(\hat{x}, e_k) - i_0(\hat{x}, e_k') \| \leq \| i_0(\hat{x}, e_k) - i_0(\hat{x}, e_k') \| \]
\[ = \left\| \prod_{i=1}^{n_0} G(\sigma_i, e_{k+1}) - \prod_{i=1}^{n_0} G(\sigma_i, e_{k+1}') \right\| \]
\[ \leq \sum_{i=1}^{n_0} \| G(\sigma_i, e_{k+1}) - G(\sigma_i, e_{k+1}') \| \]
\[ \leq \sum_{i=1}^{n_0} \frac{\delta}{3n_0} = \frac{\delta}{3}. \]

For any \( s_{i_0} \epsilon S \),
\[ \hat{a}(i(\hat{x}, e_k')) = \max \sum_{i_1, i_2, j} \{ m_{i_1} j - m_{i_2} j \}^+ = 1/2 \max \sum_{i_1, i_2, j} \{ m_{i_1} j - m_{i_2} j \} \]
\[ \leq 1/2 \max \sum_{i_1, i_2, j} \{ |m_{i_1} j - m_{i_2} j| + |m_{i_1} j - m_{i_2} j| \} \]
\[
\leq \frac{1}{2} \max_{i_1} \sum_{i_2} |m_{i_1}^j - m_{i_0}^j| + \frac{1}{2} \max_{i_2} \sum_{i_1} |m_{i_2}^j - m_{i_0}^j|
\]
\[
= \frac{1}{2} \left| |M(x, e_k') - M^0(x, e_k')|\right| + \frac{1}{2} \left| |M(x, e_k') - M^0(x, e_k')|\right|
\]
\[
= \left| |M(x, e_k') - M^0(x, e_k')|\right|.
\]

Thus, for any $s_{i_0} \in S$

\[
\exists (M(x, e_k')) \leq \left| |M(x, e_k') - M^0(x, e_k')|\right|
\]
\[
\leq \left| |M(x, e_k') - M(x, e_k)|\right|
\]
\[
+ \left| |M(x, e_k') - M^0(x, e_k')|\right|
\]
\[
\leq \delta/3 + \delta/3 + \delta/3 = \delta.
\]

But $\hat{x}$ and $k$ are arbitrary. Therefore, $\exists (M(x, e_k')) < \delta < 1$

$\forall e \in E^*(e)$ and $\forall x \in S^*$ such that $L(x) = n_0$. Let $k_i = in_0$. Hence,

by lemma 3.3.7, $A$ is ever-ergodic at $e_i$. □

**Definition 3.3.7.** Let $A = (S, \pi, G, n_0, F, E)$ be an ADE with natural metric $d^N$ and $E(e) = \{e_j|j \in J_+\}$. $G$ is uniformly continuous on $E(e)$ iff given any $\epsilon > 0$, there exists $\delta > 0$ such that

\[
\max_{\sigma \in \Sigma} |G(\sigma, e) - G(\sigma, e')| < \epsilon \text{ whenever } d^N(e, e') < \delta \text{ for any } e \in E(e).
\]

The choice of $\delta$ is independent of $e \in E(e)$.

**Corollary 3.3.5.** If $A$ is ever-ergodic at $e \in E^\infty$ and $G$ is uniformly continuous on $E(e)$, then there exists $\delta > 0$ such that for
any $e', e \in \mathcal{E}$, where $\sup_{i \in J_+} d_N(e_i, e'_i) < \delta$, $A$ is ever-ergodic at $e'$.

**Proof:** By theorem 3.3.4 there exists $\epsilon^* > 0$ such that if

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \epsilon^* \quad \forall i \in J_+$$

and $A$ is ever-ergodic at $e_i$, then $A$ is ever-ergodic at $e'_i$.

Uniform continuity of $G$ on $E(e)$ implies that given any $\epsilon > 0$, there exists $\delta > 0$ such that when $\sup_{i \in J_+} d_N(e_i, e'_i) < \delta$, we have

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \epsilon/n \quad \forall i \in J_+.$$  Hence, by lemma 3.2.3

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| \leq n \max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < n(\epsilon/n) = \epsilon \quad \forall i \in J_+.$$  Thus, if we choose $0 < \epsilon < \epsilon^*$, we can find $\delta > 0$ such that

$$\sup_{i \in J_+} d_N(e_i, e'_i) < \delta \text{ implies } \max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \epsilon^* \quad \forall i \in J_+.$$  By theorem 3.3.4 $A$ is ever-ergodic at $e'_i$.

**Theorem 3.3.6.** Let $A = (\Sigma, S, G, \pi_0, F, \mathcal{E})$ be an ADE that is ever-ergodic at $e$. For any $\epsilon > 0$, there exists $\epsilon^* > 0$ such that

$$|M(x, e) - M(x, e')| < \epsilon \quad \forall x \in \Sigma^* \quad \text{whenever} \quad \max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \epsilon^* \quad \forall i \in J_+.$$  **Proof:** By theorem 3.3.4 there exists $\epsilon_1 > 0$ such that $A$ is ever-ergodic at $e_i$ whenever

$$\max_{\sigma \in \Sigma} |G(\sigma, e_i) - G(\sigma, e'_i)| < \epsilon_1, \quad \forall i \in J_+.$$  Hence, $A$ is ever-ergodic at both $e$ and $e'$. So, given any $\epsilon > 0$, we can find $n_0$ such that $\exists (M(x, e_{k-1})) < \epsilon/6$ and $\exists (M(x, e_k)) < \epsilon/6 \quad \forall k \in J$ and $\forall x \in \Sigma^*$ such that $L(x) > n_0$. 


Choose $0 < \varepsilon_2 < \varepsilon/3n_0$. By lemma 3.2.2 we see that for any $k \in J$ \[|M(x, e_k) - M(x, e_{k_1}')| < \varepsilon/3 \] such that $L(x) \leq n_0$ whenever \[
abla \in E \] 
\\
Now let $\varepsilon^* = \min(\varepsilon_1, \varepsilon_2)$. The above results remain true when \[
abla \in E \]
\\
Since $e_0 = e$ and $e_0' = e'$, we have \[|M(x, e) - M(x, e')| < \varepsilon/3 < \varepsilon \] such that $L(x) \leq n_0$ whenever \[\nabla \in E \] \[\nabla \in E \]
\\
Thus, we have the desired result \[\forall \nabla \in E \] such that $L(x) < n_0$.
\\
Now let $x \in E$ such that $L(x) \geq n_0$. We can represent $x = M_{xy},$ where $L(y) = n_0$ and $L(x) = k \geq 0$ for some $k \in J$. Hence
\\
\[
|M(x, e) - M(x, e')| = |M(x, e)M(y, e_{k_1}) - M(x, e')M(y, e'_{k_1})|
\\
\leq |M(x, e)M(y, e_{k_1}) - M(y, e_{k_1})| + |M(x, e')M(y, e'_{k_1}) - M(y, e'_{k_1})| + |M(x, e_{k_1}) - M(y, e'_{k_1})| + 2\delta(M(y, e_{k_1})) + 2\delta(M(y, e'_{k_1}))
\\
\leq 2\delta(M(y, e_{k_1})) + 2\delta(M(y, e'_{k_1}))
\\
|\nabla \in E \]
\\
Since $L(y) = n_0$, then $\delta(M(y, e_{k_1})) < \varepsilon/6$ and $\delta(M(y, e'_{k_1})) < \varepsilon/6$. Also \[|M(y, e_{k_1}) - M(y, e'_{k_1})| < \varepsilon/3.\] Thus
\\
\[|M(x, e) - M(x, e')| < 2(\varepsilon/6) + 2(\varepsilon/6) + \varepsilon/3 = \varepsilon\]
\\
\[\forall \nabla \in E \] such that $L(x) \geq n_0$ whenever \[\nabla \in E \]
\\
Thus, the result is true \[\forall \nabla \in E \].
Theorem 3.3.6 is a useful stability result. Moreover, under certain continuity conditions we have \( s \)-stability.

Corollary 3.3.7. Let \( A = (\Sigma, S, G, \pi_0, F, E) \) be an ADE with natural metric \( d^N \) on \( E \). If for some \( e \in E^\infty \),

1. \( A \) is ever-ergodic at \( e \)
2. \( G \) is uniformly continuous on \( E(e) = \{ e_i | i \in I_+ \} \)
3. \( \pi_0 \) is continuous at \( e_0 \),

then \( A \) is \( s \)-stable at \( e \).

Proof: By theorem 3.3.6, given any \( \varepsilon > 0 \), there exists \( \varepsilon^* > 0 \) such that

\[
| |H(x, e) - M(x, e^i)| | < \varepsilon/2 \quad \forall x, e \in \Sigma^*
\]

whenever

\[
\max_{\sigma \in \Sigma} | G(\sigma, e_1) - G(\sigma, e_1^i) | | < \varepsilon^* \quad \forall \sigma \in I_+.
\]

Since \( G \) is uniformly continuous on \( E(e) \), then there exists \( \delta_1 > 0 \) such that

\[
\max_{\sigma \in \Sigma} | G(\sigma, e_1) - G(\sigma, e_1^i) | | < \varepsilon^* \quad \forall \sigma \in I_+ \text{ whenever } \sup_{i \in I_+} d^N(e_1, e_1^i) < \delta_1.
\]

The initial distribution function \( \pi_0 \) is continuous at \( e_0 \), so there exists \( \delta_2 > 0 \) such that

\[
| |\pi_0(e_0) - \pi_0(e_0^i)| | < \varepsilon/2n \quad \text{when } d^N(e_0, e_0^i) < \delta_2.
\]

Hence by lemma 3.2.3

\[
| |\pi_0(e_0) - \pi_0(e_0^i)| | < \varepsilon/2 \quad \text{when } d^N(e_0, e_0^i) < \delta_2.
\]

Choose \( \delta = \min(\delta_1, \delta_2) \). Hence, whenever

\[
\sup_{i \in I} d^N(e_1, e_1^i) < \delta,
\]

\[
| |\pi(x, e) - \pi(x, e^i)| | \leq | |\pi(x, e) - \pi(x, e^i)| |
\]

by lemma 3.2.3

\[
= | |\pi_0(e_0)H(x, e) - \pi_0(e_0^i)H(x, e)| |
\]

by lemma 3.2.1

\[
\leq | |\pi_0(e_0) - \pi_0(e_0^i)| | + | |H(x, e) - H(x, e^i)| |
\]

\[
< \varepsilon/2 + \varepsilon/2 = \varepsilon \quad \forall x \in \Sigma^*.
\]

Hence, \( A \) is \( s \)-stable at \( e \). \( \square \)
Example 3.3.1. Let \( A = (\Sigma, S, G, \pi_0, F, E) \) be an ADE with 
\( \Sigma = \{0\}, S = \{s_1, s_2, s_3\} \) and \( E = [-1/2, 1/2] \). Let Euclidean distance
be the natural metric on \( E \); that is, for any \( e, e' \in E \), \( d^N(e, e') = |e-e'| \).
The function \( G \) is defined
\[
G(\sigma, e) = \begin{pmatrix}
1/3+e^5 & 0 & 2/3-e^5 \\
e^{4/2} & e^{4/2} & 1-e^4 \\
(3+e^5)/5 & (2-e^5)/5 & 0
\end{pmatrix}
\]
We define \( \pi_0(e) = (1/6+e^4, 1/3+e^4, 1/2-2e^4) \).
Let \( e \in E^\infty \) such that \( \liminf_{i \to \infty} |e_i| \geq \gamma \) for some \( \gamma > 0 \). Hence, there exists \( N \) such that \( |e_i| \geq \gamma/2 > 0 \) \( \forall i \geq N \). But \( G(\sigma, e) \)
is \( \gamma^4/32 \)-scrambling \( \forall e \in E \) such that \( |e| \geq \gamma/2 \). Hence, for
\( \mathbb{N}(\gamma^4/32) = 1 \), \( A \) satisfies the ever-\( H_4(\gamma^4/32) \) condition at \( e_N \).
By theorem 3.3.3 \( A \) is ever-ergodic at \( e \).

The components of \( G \) are continuous at each \( e \in E \). But \( E \) is
a compact set with respect to the natural metric topology. Hence, \( G \) is
uniformly continuous. Clearly, \( G \) is uniformly continuous on
\( E(e) \) for any \( e \in E^\infty \). The function \( \pi_0 \) is also continuous at any \( e \in E \).

Hence, given any \( \epsilon > 0 \), we can find \( \delta > 0 \) such that
\[
|\pi(x, e) - \pi(x, e')| < \epsilon \quad \forall x \in \Sigma \] whenever \( \sup_{i \in J} d^N(e_i, e'_i) = \sup_{i \in J} |e_i - e'_i| < \delta \).

Let \( A = (\Sigma, S, G, \pi_0, F, E) \) be an ADE with natural metric \( d^N \) on \( E \).
Let \( E \) be a compact set with respect to the natural metric topology.
If the components of \( G(\sigma, e) \) are continuous at each \( e \in E \), then \( G \) is
uniformly continuous with respect to \( d^N \). Moreover, if there exists
\( \delta > 0 \) such that all of the components of \( G(\sigma, e) \) are bounded below by
γ, then A is w-ergodic and, hence, ever-ergodic at each e ∈ E∞. An ADE satisfying the above conditions is s-stable at every e ∈ E∞. Furthermore, the choice of δ need not depend on e.

3.4 Response Intervals and a-Stability

Now we shall establish sufficient conditions for an ADE to be tape-acceptance stable under environment sequence e = (e_0, e_1, ...). We shall consider the problem when we have that λ is an isolated cut-point for A under total environment e. Recall that a cut-point λ for A under total environment e is isolated if there exists γ > 0 such that |p(x, e) - λ| ≥ γ ∀x ∈ E*. In view of theorem 3.1.2 and corollary 3.3.7, we shall not require the automaton to be s-stable or ever-ergodic at e. We shall tolerate perturbations of the transition functions as long as the acceptance probability function does not cross the cut-point.

Let us consider an ADE A = (Σ, S, G, π_0, F, E) and some total environment e ∈ E∞. Define the response intervals R_i, i = 1, 2, ..., as

R_i = [0, 1] ∩ \{ u \mid \sum_{σ ∈ Σ} Λ(σ, i, j), \sum_{j ∈ F} Ψ(σ, i, j) \}

where Λ(σ, i, j) = \min_k g_{kj}(σ, e_1^i) and Ψ(σ, i, j) = \max_k g_{kj}(σ, e_1^i).

The values g_{kj}(σ, e_1^i) are the components of G(σ, e_1^i). We also define

R_0 = \{ p(A, e) \} = \{ π_0(e_0) η^F \}.

Theorem 3.4.1. Let A = (Σ, S, G, π_0, F, E) be an ADE. For any e ∈ E∞ and λ ∈ (0, 1), T(A, e, λ) = T(A, e_i, λ) whenever
1. \[ \lambda \not\vDash \bigvee_{i \in J} R_i \, . \]

2. \[ \inf_{r_i \in R_i} \frac{|\lambda - r_i|}{\#(F)} > |G(\sigma, e_1) - G(\sigma, e_1')| \quad \forall \sigma \in \Sigma \text{ and } \forall i \in J_+ . \]

3. \[ |\lambda - p(\lambda, e) - p(\lambda, e')| > |p(\lambda, e) - p(\lambda, e')| , \]

where \( R_i, \ i=0,1, \ldots, \) are defined for \( A \) under total environment \( e \) as above.

**Proof:** Consider \( \Lambda \), the empty tape. \( R_0 = \{ p(\lambda, e) \} \).

\( \lambda \not\vDash \bigvee_{i \in J} R_i \) implies \( \lambda \not\vDash p(\lambda, e) \).

If \( \Lambda \in T(\lambda, e, \lambda) \), then \( p(\lambda, e) > \lambda \geq 0 \). Now \( p(\lambda, e') = p(\lambda, e) + p(\lambda, e') - p(\lambda, e) \leq |p(\lambda, e) - p(\lambda, e')| \). Hence, by condition 3, \( p(\lambda, e') > p(\lambda, e) - |\lambda - p(\lambda, e)| \geq \lambda \). Thus, \( p(\lambda, e') > \lambda \); consequently \( \Lambda \in T(\lambda, e', \lambda) \).

If \( \Lambda \not\vDash T(\lambda, e, \lambda) \), then \( 0 \leq p(\lambda, e) < \lambda \). Now

\[
p(\lambda, e') = p(\lambda, e) + p(\lambda, e') - p(\lambda, e) \\
\leq p(\lambda, e) + |p(\lambda, e) - p(\lambda, e')| \\
< p(\lambda, e) + |\lambda - p(\lambda, e)| \quad \text{by condition 3} \\
\leq \lambda .
\]

Thus \( p(\lambda, e') \leq \lambda \); consequently \( \Lambda \not\vDash T(\lambda, e', \lambda) \).

Now consider any other element of \( \Sigma^* \). We can write it in the form \( x_\sigma e \in \Sigma^* \), where \( x \in \Sigma^* \) and \( e \in \Sigma \). Since \( e \in \Sigma \), then it is impossible for \( x \sigma = \lambda \).

Let \( v \) and \( w \) be arbitrary vectors in \( V_n \). The functions \( \Lambda \) and \( \nabla \) are defined so that for any \( v \in V_n \) and \( i = L(v) + 1 \), we have
\[
\sum_{j \in F} \Delta(\sigma, i, j) \leq w_G(\sigma, e_i) \eta^F \leq \sum_{j \in F} \nu(\sigma, i, j).
\]

Condition 2 implies that for any \( w \in V_n \) that

\[
\sum_{j \in F} \Delta(\sigma, i, j) - \delta_1^F(F) \leq w_G(\sigma, e_i') \eta^F \leq \sum_{j \in F} \nu(\sigma, i, j) + \delta_1^F(F),
\]

where

\[
\delta_1 = \inf_{r_i \in R_i} \frac{|\lambda - r_i|}{\beta(F)} > 0 \quad \forall \lambda \in J_+.
\]

Now let \( v = \pi(x, e) \) and \( w = \pi(x, e') \). So \( w_G(\sigma, e_i) \eta^F = p(x_\sigma, e) \) and \( w_G(\sigma, e_i') \eta^F = p(x_\sigma, e') \).

Since \( \lambda \notin R_i \), then \( p(x_\sigma, e) \neq \lambda \). If \( x_\sigma e T(A, e, \lambda) \), then \( p(x_\sigma, e) > \lambda \). But by condition 2, \( \lambda < \sum_{j \in F} \Delta(\sigma, i, j) - \delta_1^F(F) \leq p(x_\sigma, e') \).

Hence, \( p(x_\sigma, e') > \lambda \); therefore \( x_\sigma e T(A, e', \lambda) \).

If \( x_\sigma \notin T(A, e, \lambda) \), then \( p(x_\sigma, e) < \lambda \). Again by condition 2, \( p(x_\sigma, e) \leq \sum_{j \in F} \nu(\sigma, i, j) + \delta_1^F(F) < \lambda \). So \( p(x_\sigma, e') < \lambda \); hence, \( x_\sigma \notin T(A, e', \lambda) \).

Therefore, \( x_\sigma e T(A, e, \lambda) \) iff \( x_\sigma e T(A, e', \lambda) \). But \( x_\sigma \) is any arbitrary, nonempty tape. Together with the considerations of the empty tape, we have \( T(A, e, \lambda) = T(A, e', \lambda) \). \( \square \)

**Corollary 3.4.2.** Let \( A = (\Sigma, S, G, \pi_0, F, E) \) be an ADE with natural metric \( d^N \) on \( E \). If for some \( e \in e^\infty \) and \( \lambda \in [0, 1) \),

1. \( \lambda \notin \bigcup_{i \in J_+} R_i \) (—— denotes closure),
2. \( G \) is uniformly continuous on \( E(e) \),
3. \( \pi_0 \) is continuous at \( e_0 \),

then \( A \) with cut-point \( \lambda \) is a-stable at \( e \).
Proof: We are interested in finding $\delta > 0$ such that
\[ \sup_{i \in J} d_N(e_i, e_i') < \delta \] implies the conditions of theorem 3.4.1 hold.
Clearly $\lambda \notin \bigcup_{i \in J} R_i$ implies $\lambda \notin \bigcup_{i \in J} R_i$. Also, $\lambda \notin \bigcup_{i \in J} R_i$ implies there exists $\gamma_1 > 0$ such that $\inf_{i \in J_+} \inf_{r_i \in R_i} |\lambda - r_i| = \gamma_1$. Now let $\epsilon_1 = \gamma_1^\#(F)$. By the uniform continuity of $G$ on $E(e)$, there exist $\delta_1 > 0$ such that
\[ \max_{\sigma \in \Sigma} |G(\sigma, e_i') - G(\sigma, e_i')| < \epsilon_1 = \gamma_1^\#(F) \leq \inf_{r_i \in R_i} \frac{|\lambda - r_i|}{\#(F)}, \]
for $\forall i \in J_+$ whenever $\sup_{i \in J_+} d_N(e_i, e_i') < \delta_1$.

Again, $\lambda \notin \bigcup_{i \in J} R_i$ implies $\lambda R_0$. Hence $\lambda \neq p(\Lambda, e)$. So let $\gamma_2 = |\lambda - p(\Lambda, e)| > 0$. Let $\epsilon_2 = \gamma_2^\#(F)$. By the continuity of $\pi_0$ at $e_0$, there exists $\delta_2 > 0$ such that $|\pi_0(e_0) - \pi_0(e_0')| < \epsilon_2$ whenever $d_N(e_0, e_0') < \delta_2$. Therefore
\[ |p(\Lambda, e) - p(\Lambda, e)| \leq |\pi_0(e_0) - \pi_0(e_0')|^\#(F) < \epsilon_2^\#(F) = \gamma_2 \]
whenever $d_N(e_0, e_0') < \delta_2$.

Now choose $\delta = \min(\delta_1, \delta_2)$. The conditions of theorem 3.4.1 are verified. Hence, there exists $\delta > 0$ such that $\Psi_{e'} eT_\infty$, where $\sup_{i \in J} d_N(e_i, e_i') < \delta$, then $T(A, e, \lambda) = T(A, e', \lambda)$. That is, $A$ with cut-point $\lambda$ is a-stable at $e$.

Corollary 3.4.3. Let $A = (\Sigma, S, G, \pi_0, F, E)$ be an ADE. For any $e \in T_\infty$, $\lambda$ is an isolated cut-point for $A$ under total environment $e$. 

\[ \Box \]
if $\lambda \notin \overline{\cup_{i \in J} R_i}$.

**Proof:** Clearly, $\{p(x,e)| x \in L^*\} \subseteq \overline{\cup_{i \in J} R_i}$. But $\lambda \notin \overline{\cup_{i \in J} R_i}$ and $\overline{\cup_{i \in J} R_i}$ is a closed set. Hence, there exists $\gamma > 0$ such that $|p(x,e) - \lambda| \geq \gamma \forall x \in L^*$. \qed

**Example 3.4.1.** Let $\mathcal{A} = (\Sigma, S, G, \pi_0, F, E)$ be an ADE with $E = \{\sigma_1, \sigma_2\}$, $S = \{s_1, s_2, s_3, s_4\}$, $F = \{s_2, s_3\}$, and $E = [-1/2, 1/2]$. Let Euclidean distance be the natural metric on $E$. We define $G$ for $\sigma_1$ as

$$G(\sigma_1, e) = \begin{pmatrix}
11/12 - e^4 & 1/12 + e^4 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1/6 + e^4 & 5/6 - e^4
\end{pmatrix}$$

and for $\sigma_2$ as

$$G(\sigma_2, e) = \begin{pmatrix}
1/12 & 1/12 - e^4 & 5/6 + e^4 & 0 \\
0 & 1/4 + e^4 & 3/4 - e^4 & 0 \\
0 & 0 & 1 & 0 \\
2e^4 & 0 & 3/4 - 2e^4 & 1/4
\end{pmatrix}.$$  

Define $\pi_0(e) = (e_4, 1/4 - 3e^4, e^4, 3/4 + e^4)$.

$E$ is a compact set with respect to the natural metric topology. Since for each $\sigma \in \Sigma$ the components of $G(\sigma, e)$ are continuous at each $e \in E$, then $G$ is uniformly continuous with respect to $d^N$. For any $e \in E$, $\pi_0$ is also continuous at $e$ with respect to $d^N$.  


Let \( e = (e_0, e_1, \ldots) \in \mathbb{E}^\infty \). Thus \( \mathbb{R}_0 = \{1/4 - 2e_0^4\} \) and 
\[ \mathbb{R}_1 = [0,1/4 + 2e_1^4] \cup [3/4 - 2e_1^4,1]. \]
Clearly, \( \bigcup_{i \in J} \mathbb{R}_i \subset [0,3/8] \cup [5/8,1] \). Hence, for any \( \lambda \in (3/8,5/8) \), \( \mathbb{A} \) with cut-point \( \lambda \) is a-stable at any \( e \in \mathbb{E}^\infty \). Note that any \( \lambda \in (3/8,5/8) \) is an isolated cut-point.

For any \( \delta > 0 \) and any \( e \in \mathbb{E}^\infty \), there does not exist an integer \( k > 0 \) such that \( \mathbb{M}(\mathbb{A}, e) = \prod_{i=1}^{k} \mathbb{G}(\sigma_i, e_i) \) is \( \delta \)-scrambling. Hence, \( \mathbb{A} \) does not satisfy the ever-\( \mathcal{H}_4(\delta) \) condition at any \( e \in \mathbb{E}^\infty \) for any \( \delta > 0 \). Thus, \( \mathbb{A} \) is not ever-ergodic at any \( e \in \mathbb{E}^\infty \).

3.5 Round Off Considerations

Suppose we have an ADE \( \mathbb{A} = (\mathbb{E}, \mathbb{S}, \mathbb{G}, \pi_0, \mathbb{F}, \mathbb{E}) \) with natural metric \( d^N \) on \( \mathbb{E} \). We, however, are not able to measure the environment configuration exactly; the environment can only be measured correctly up to a certain approximation and the remainder is subject to round off.

That is, if \( e \in \mathbb{E} \) is the true condition of the environment, we can only observe a unique round off estimate \( \mathbb{f}_\gamma(e) \) such that \( d^N(e, \mathbb{f}_\gamma(e)) \leq \gamma \) for some \( \gamma > 0 \). Thus, we shall assume that we have a subset 
\[ F_\gamma = \{ \mathbb{f}_\gamma(e) | e \in \mathbb{E} \} \] that is \( \gamma \)-dense in \( \mathbb{E} \). That is, for any \( e \in \mathbb{E} \), there exists a unique \( f \in F_\gamma \) such that \( d^N(e,f) \leq \gamma \).

For notational convenience, we shall let \( \mathbb{f}_\gamma(e) \) denote the component-wise rounded value of \( e \). Thus \( \mathbb{f}_\gamma(e) = (f_\gamma(e_0), f_\gamma(e_1), \ldots) \).

Let \( \mathbb{F}_\gamma^\infty = \{ (f_0, f_1, \ldots) | f_i \in F_\gamma \} \). Clearly, \( \mathbb{F}_\gamma^\infty = \{ \mathbb{f}_\gamma(e) | e \in \mathbb{E}^\infty \} \). For any \( e \in \mathbb{E}^\infty \), \( \sup_{i \in J} d^N(e_i, f(e_i)) \leq \gamma \). Hence, \( \mathbb{F}_\gamma^\infty \) is \( \gamma \)-dense in \( \mathbb{E}^\infty \) with respect to the metric \( \sup_{i \in J} d^N \).
Now we can use the stability results of the previous sections to find round off schemes which maintain stability of an acceptance set or stability of the acceptance probability function.

Suppose we have an ADE \( A \) which is \( a \)-stable at some cut-point \( \lambda \) for some environment sequence \( \underline{f}_\infty \). There exists \( \delta > 0 \) such that \( T(A, e, \lambda) = T(A, f, \lambda) \) whenever \( \sup_{i \in J} d^N(e, e_i) < \delta \). Clearly, if \( \gamma < \delta \), then \( \forall \varepsilon \in \mathbb{E}^\infty \) such that \( f(\varepsilon) = f \), we have \( \sup_{i \in J} d^N(f, e_i) = \sup_{i \in J} d^N(f(\varepsilon), e_i) \leq \gamma < \delta \). Thus for sufficiently small \( \gamma > 0 \),

\[
T(A, f, \lambda) = T(A, e, \lambda) \quad \forall \varepsilon \in \mathbb{E}^\infty \text{ such that } f(\varepsilon) = f.
\]

So if we have \( a \)-stability for some cut-point \( \lambda \) at the rounded value of the environment sequence and if the round off error is sufficiently small, then the set defined by the ADE with cut-point \( \lambda \) remains unchanged by the round off procedure. We also obtain this result under the conditions of corollary 3.4.2 when \( \gamma \) is sufficiently small.

Now suppose that \( A \) is \( s \)-stable at some \( \underline{f}_\infty \). Hence, given any \( \varepsilon > 0 \), there exists \( \delta(\varepsilon) \) such that \( |\pi(x, f) - \pi(x, e)| < \varepsilon \)

\( \forall x \in \mathbb{E}^* \) whenever \( \sup_{i \in J} d^N(f, e_i) < \delta(\varepsilon) \). Clearly, if \( \gamma < \delta(\varepsilon) \), then \( \forall \varepsilon \in \mathbb{E}^\infty \) such that \( f(\varepsilon) = f \), we have \( |\pi(x, f_{\gamma}(\varepsilon)) - \pi(x, e)| < \varepsilon \quad \forall x \in \mathbb{E}^* \).

For \( \gamma \) sufficiently small, the conditions of corollary 3.3.7 also imply this result.

The round off procedure is also applicable to approximation of the acceptance probability function for sets of tapes of bounded length. In this case we need only consider \( f_0, f_1, \ldots, f_k \), for some \( k \in J \). For any \( \varepsilon > 0 \), by theorem 3.2.1 if \( \gamma \) is sufficiently small, then

\[
|p(x, e) - p(x, f)| < \varepsilon \quad \forall x \in \mathbb{E}^* \text{ such that } L(x) \leq k, \text{ whenever } f_{\gamma}(e_i) = f_i \text{ for } i = 1, \ldots, k.
\]
CHAPTER IV

Automata in Random Environments

4.1. Basic Notions.

We shall now consider the case of automata operating within random environments. In contrast to the case of the ADE, where we suppose we can control the environment, we shall assume that the realization of the environment is governed by some probabilistic structure. This formulation is useful when we are only able to make certain probabilistic assumptions about the occurrence of any environment configuration or when the environment configuration can only be measured by statistical techniques.

Let \( A = (\Sigma, S, C, \pi_0, F, E) \) be an automaton in a deterministic environment. Let \((\Omega, \mathcal{B}, P)\) be a probability space. For each \( j \in \{0, 1, \ldots \} = J \), let \( z_j \) be a measurable function from \((\Omega, \mathcal{B})\) to \((E, E')\), where \( E' \) is a \( \sigma \)-field of subsets of \( E \). We shall use \( z_0 \) to denote the random variable for the initial configuration of the environment and \( z_j, j=1,2,\ldots \), to denote the random variable for the configuration of the environment for the \( j \)-th input symbol. The family of random variables \( Z = \{z_j | j \in J\} \) is an environmental stochastic process (ESP). When we wish to emphasize the probability structure, we shall use \((\Omega, \mathcal{B}, P, Z)\) to denote the process.
Definition 4.1.1. An automaton in a random environment (ARE) is a system \((A, Z)\), where \(A\) is an ADE and \(Z\) is an ESP.

Let \(A = (\Sigma, S, G, \pi_0, F, E)\) be an ADE and \((\Omega, B, P, Z)\) be an ESP. Note that \(G : \Sigma^* E \rightarrow M^\infty\) and \(z_j : \Omega \rightarrow E\) for all \(j \in J^+ = \{1, 2, \ldots\}\). Now for all \(\sigma \in \Sigma\), we define \(G_\sigma(e) = G(\sigma, e)\) and assume \(G_\sigma(e)\) is measurable. Thus, \(G_\sigma \cdot z_j : \Omega \rightarrow M^\infty\) such that \(G_\sigma \cdot z_j(\omega) = G(\sigma, z_j(\omega))\) is measurable. For convenience we shall call \(G_\sigma \cdot z_j\) by the name \(G(\sigma, z_j)\). Similarly, \(z_0 : \Omega \rightarrow E\) and we assume that \(\pi_0 : E^\infty \rightarrow V^\infty\) is also measurable. Hence, \(\pi_0 \cdot z_0 : \Omega \rightarrow V^\infty\) such that \(\pi_0 \cdot z_0(\omega) = \pi_0(z_0(\omega))\) is measurable. We shall call \(\pi_0 \cdot z_0\) by the name \(\pi_0(z_0)\). Note that \(G(\sigma, z_j)\) for all \(j \in J^+\) and all \(\sigma \in \Sigma\) is a random stochastic matrix and \(\pi_0(z_0)\) is a random stochastic vector. We shall, however, encounter no ambiguity in using the notation \(G\) and \(\pi_0\) in both the ADE and ARE contexts.

Let \((\Omega, \mathcal{B}, P, Z)\) be an ESP with environment set \(E\). Define \(z : \Omega \rightarrow E^\infty\) by \(z(\omega) = (z_0(\omega), z_1(\omega), \ldots)\). The mapping \(z\) is measurable map from \((\Omega, \mathcal{B})\) to \((E^\infty, B^\infty)\), where \(B^\infty\) is the smallest \(\sigma\)-field generated by the measurable cylinders. We can use this to define the random matrix transition function. Let \(x \in \Sigma^*\) such that \(L(x) = k\). Now, \(H(\sigma, z(\omega)) = \prod_{i=1}^{k} G_{\sigma_i} \cdot z_i(\omega)\). Similarly, we have \(\pi(x, z(\omega)) = \pi_0(z_0(\omega)) H(\sigma, z(\omega))\) and \(p(x, z(\omega)) = \pi(x, z(\omega)) \eta^P\). Furthermore, \(T(A, z, \lambda)\) for any \(\lambda \in [0, 1]\) is a set function of \(\omega \in \Omega\) taking values in the set of all subsets of \(\Sigma^*\).

The functions \(\pi_0(z_0)\) and \(\pi(x, z)\) for any \(x \in \Sigma^*\) take values in \(V^\infty\). It is not difficult to see that \(V^\infty\) is a convex set; thus, \(E\pi_0(z_0) \in V^\infty\) and \(E\pi(x, z) \in V^\infty\) for any \(x \in \Sigma^*\). Similarly, \(M^\infty\) is a convex set; hence \(E G(\sigma, z_j) \in M^\infty\) for all \(\sigma \in \Sigma\) and for all \(j \in J^+\) and
$E(x,z) \in \mathcal{M}_n$ for all $x \in \mathcal{E}$. Note also that $[0,1]$ is a convex set; so $E(x,z) \in [0,1]$ for all $x \in \mathcal{E}$.

**Definition 4.1.2.** Let $(\mathcal{A}, Z)$ be an ADE and $\lambda$ a real number, $\lambda \in [0,1)$. The set of tapes $E(T(\mathcal{A}, Z, \lambda))$ defined by $E(T(\mathcal{A}, Z, \lambda)) = \{x | x \in \mathcal{E}^*, \lambda < E(p(x, z))\}$ is called the expected set of tapes accepted, or defined, by $(\mathcal{A}, Z)$ with cut-point $\lambda$.

Suppose $z(1), z(2), \ldots$ are independent identically distributed (IID) sequences of random environments. Thus, for any fixed, but arbitrary $x \in \mathcal{E}^*, \{p(x, z^{(i)})\}_{i=1}^\infty$ is a sequence of IID random variables since $|p(x, z^{(i)})| \leq 1$ a.s. for all $i=1, 2, \ldots$, we may apply the strong law of large numbers to obtain

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N p(x, z^{(i)}) = E(p(x, z)) \text{ a.s.}$$

We can obtain similar almost sure convergence for the components of $\pi$ and $\mathcal{M}$.

Let $A = (\mathcal{E}, S, G, \pi_0, F, E)$ be an ADE. Let us consider the probability space $([0,1], \mathcal{B}([0,1]), \mu)$, where $\mathcal{B}([0,1])$ is the relative Borel field and $\mu$ is Lebesgue measure. We define the random variable $\psi(\theta) = \theta$, $\theta \in [0,1]$, on this probability space. For any fixed, but arbitrary $x \in \mathcal{E}^*$ and $\theta \in \mathcal{E}$, we define the random variables

$$I(x, z, \psi(\theta)) = \begin{cases} 1 \text{ if } \psi(\theta) \leq p(x, z) \\ 0 \text{ otherwise} \end{cases}.$$
Clearly, for any $x \in \mathcal{E}^*$ and $e \in \mathcal{E}^\omega$, $I(x,e,\psi)$ is measurable and
\[ u(\theta | I(x,e,\psi(\theta)) = 1) = p(x,e). \]
Also, $E \ I(x,e,\psi) = \int I(x,e,\psi(\theta)) d\mu(\theta) = p(x,e)$. The random variable $I(x,e,\psi)$ has the same relevant probability structure as an indicator which is 1 when $A$ is in a state in $F$ after input $x$ under total environment $e$ and 0 otherwise.

Let $(\Omega, \mathcal{E}, P, Z)$ be an ESP with environment set $E$. We have defined $z: \Omega \rightarrow \mathcal{E}^\omega$. Hence, we can consider the composition $I(x,z(\cdot),\psi(\cdot))$ as a function defined on the product space $[0,1] \times \Omega$, where
\[ I(x,z(\omega),\psi(\theta)) = \begin{cases} 1 & \text{if } \psi(\theta) \leq p(x,z(\omega)) \\ 0 & \text{otherwise} \end{cases}. \]
So for any fixed, but arbitrary $x \in \mathcal{E}^*$, $I(x,z,\psi)$ is a random variable defined on the product probability space. We now obtain $E \ I(x,z,\psi)$ with respect to the product measure.

\[ E \ I(x,z,\psi) = \int_{[0,1] \times \Omega} I(x,z(\omega),\psi(\theta)) d(\mu \times P). \]

By Fubini's theorem,
\[ E \ I(x,z,\psi) = \int_{[0,1]} \left( \int_{\Omega} I(x,z(\omega),\psi(\theta)) d\mu(\theta) dp(\omega) \right). \]
\[ = \int_{\Omega} p(x,z(\omega)) dp(\omega) = E p(x,z), \]
where the last expectation is taken with respect to $(\Omega, \mathcal{E}, P)$.

Suppose $\{(z^{(i)},\psi^{(i)})\}_{i=1}^\infty$ are independent copies of $(z,\psi)$. Hence $\{I(x,z^{(i)},\psi^{(i)})\}_{i=1}^\infty$ is a sequence of random variables which are IID according to the product measure $\mu \times P$. By the strong law of large numbers,
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} I(x, z(1), \psi(1)) = E \ p(x, z) \ a.s. \ \mu \times P. \quad \text{Thus, if we repeatedly}
\]
input tape \( x \) under IID random environment sequences, the relative frequency of acceptance of \( x \) is a strongly consistent, unbiased estimator of \( E \ p(x, z) \).

4.2 Random Environment Sequences of Independent Random Variables

Let \((A, Z)\) be an ARE such that \( z = (z_0, z_1, \ldots) \) is a sequence of independent random variables. Thus, for any \( x \in \mathcal{E}^* \) such that

\[
L(x) = k, \ E_p(x, z) = E\{\pi_0(z_0) \left( \prod_{i=1}^{k} G(\sigma_1, z_1) \right)^{\eta} \} = (E\pi_0(z_0)) \left( \prod_{i=1}^{k} E G(\sigma_1, z_1) \right)^{\eta}
\]

And, similarly, the expectations of the other automata functions are products of the expectations of their factors.

Let \( Z(z) = \{z_j | j \in J\} \) be the set of all mappings in the sequence \( z \).

For our purposes here, we shall consider two environmental random variables to be equal if their distribution functions are equal. Let \( E' \) be a set which is in one-to-one correspondence with \( Z(z) \). That is, we have a functional \( g: Z(z) \to E' \), which is a one-to-one correspondence. For any \( z \in Z(z) \) and \( \sigma \in \mathcal{E} \), define \( \pi_0'(g(z)) = E\pi_0(z) \) and \( G'(\sigma, g(z)) = E G(\sigma, z) \). Thus, \( A' = (\mathcal{E}, \mathcal{S}, G', \pi_0', F, E') \) is an ADE. Also, for any \( x \in \mathcal{E}^* \),

\[
p'(x, g(z)) = \pi_0'(g(z_0)) \left( \prod_{i=1}^{k} G(\sigma_1, g(z_1)) \right)^{\eta} = (E\pi_0(z_0)) \left( \prod_{i=1}^{k} E G(\sigma_1, z_1) \right)^{\eta} = E_p(x, z).
\]
Therefore, for any $\lambda \epsilon [0, 1)$, $T(A^1, g(z), \lambda) = ET(A, z, \lambda)$. Furthermore, if $Z(z)$ contains only a finite number of random variables with distinct distributions, then $E'$ is a finite set. Hence, $A'$ can be simulated by a PA. In the particular case when $z$ is a sequence of IID random variables, then $\#(E') = 1$ and we can find a PA $A''$ over $\Sigma$ such that $T(A'', \lambda) = ET(A, z, \lambda)$ for all $\lambda \epsilon [0, 1)$.

Suppose we have an ARE $(A, Z)$ where $\Sigma = \{\sigma\}$ and $z_1, z_2, \ldots$ is a sequence of IID random variables. The following results hold for state sets of $A$ of any finite cardinality, but for the sake of simplicity of exposition we shall restrict consideration to $S = \{s_1, s_2\}$. Let $\sigma^N \epsilon \Sigma^*$ denote the tape of $N$ successive inputs of $\sigma$. We wish to find the asymptotic moments of $M(\sigma^N, z)$. Let $(\alpha_{N, \beta_{N}})$ be the first row of $M(\sigma^N, z)$. Clearly, for every $\omega \epsilon \Omega$, $(\alpha_{N, \beta_{N}}) \epsilon V_2$ for all $N = 1, 2, \ldots$. Clearly, $(\alpha_{N+1, \beta_{N+1}}) = (\alpha_{N, \beta_{N}}) G(\sigma, z_{N+1})$. By the IID property,

$$(E_{N+1} \alpha, E_{N+1} \beta) = (E_{N} \alpha, E_{N} \beta) EG(\sigma, z_{N+1}) = (E_{1} \alpha, E_{1} \beta) (EG(\sigma, z_{1}))^N.$$  

We say a matrix $A \epsilon M_n$ is ergodic if it can be considered as a transition matrix of an irreducible ergodic Markov chain as defined in Feller [6].

Thus, if $EG(\sigma, z_1)$ is an ergodic matrix, then $(EG(\sigma, z_1))^N$ converges to a limit matrix $L_1$. The rows of $L_1$ are identical; thus $(E_{N} \alpha, E_{N} \beta)$ converges to $(E_{1} \alpha, E_{1} \beta) L_1 = \ell_1$, where $\ell_1$ is any row of $L_1$.

Let $k$ be any fixed positive integer. For every $\omega \epsilon \Omega$ and $N = 1, 2, \ldots$, $(\alpha_{N, \beta_{N}})^k = 1$. Thus $(\alpha_{N+1, \beta_{N+1}} \epsilon V_{k+1}$ for every $\omega \epsilon \Omega$ and $N = 1, 2, \ldots$. From the above, we know that

$$\begin{align*}
\alpha_{N+1} &= \alpha_{N} \beta_{11}(\sigma, z_{N+1}) + \beta_{N} \beta_{21}(\sigma, z_{N+1}) \\
\beta_{N+1} &= \alpha_{N} \beta_{12}(\sigma, z_{N+1}) + \beta_{N} \beta_{22}(\sigma, z_{N+1})
\end{align*}$$
where $g_{ij}^{(s)}(z_{N+1}^*)$ is the $(i,j)$ component of $G(s,z_{N+1}^*)$. By taking
the appropriate products we can derive an expression for each component
of $(a_{N+1}^{k}, a_{N}^{k-1}b_{N+1}^N, ..., b_{N+1}^k)$ in terms of the components of
$(a_{N}^{k}, a_{N}^{k-1}b_{N}^N, ..., b_{N}^k)$ and $G(s,z_{N+1}^*)$. This reduces to the relation
$(a_{N+1}^{k}, a_{N+1}^{k-1}b_{N+1}^N, ..., b_{N+1}^k) = (a_{N}^{k}, a_{N}^{k-1}b_{N}^N, ..., b_{N}^k)\cdot G(s,z_{N+1}^*)[k],$
where $G(s,z_{N+1}^*)[k]$ is the $k$-th Kronecker power of $G(s,z_{N+1}^*)$ as
defined in Bellman [5]. Observe that $G(s,z_{N+1}^*)[k] \in M_{k+1}^k$. Again, by the
IID property
\[
(\mathbb{E}^k_{N+1}, \mathbb{E}^k_1 a_{N+1}^{k-1}b_{N+1}^N, ..., b_{N+1}^k) = (\mathbb{E}^k_N, \mathbb{E}^k_1 a_{N}^{k-1}b_{N}^N, ..., b_{N}^k)(\mathbb{E}(s,z_{N+1}^*)[k])^N.
\]
If $\mathbb{E}(s,z_{1}^*)[k]$ is an ergodic matrix, $(\mathbb{E}(s,z_{1}^*)[k])^N$ converges to a
limit $L_k$ with identical rows. Thus as before,
$(\mathbb{E}^k_N, \mathbb{E}^k_1 a_{N}^{k-1}b_{N}^N, ..., b_{N}^k)$ converges to $\ell_k$, a row of $L_k$. So $\ell_k$
is a vector of asymptotic moments of the first row of $M(s^N,a)$.

4.3 Limit Results for Sequences of Environment Processes

Let $A = (\Sigma, S, G, \pi_0, F, \mathbb{E})$ be an ADE and $(\Omega, B, P)$ be a probability
space. Let $\ell_\alpha(\Omega, B, P; \mathbb{E})$ be the space of all environment valued sequences
defined on $\Omega$ such that for any
\[
z = (z_0, z_1, ..., \in \ell_\alpha(\Omega, B, P; \mathbb{E}), \sum_{j \in J} E|z_j(\omega)|^\alpha < \infty
\]
for almost every $\omega \in \Omega$. For the sake of generality, we do not give an
explicit expression for the norm of the environment configuration.
For \( z^{(1)}, z^{(2)} \in \ell^e_\alpha(\Omega, B, P; \mathbb{E}) \), if we consider \( z^{(1)} = z^{(2)} \) iff \( z^{(1)}(\omega) = z^{(2)}(\omega) \) a.s., then \( \ell^e_\alpha(\Omega, B, P; \mathbb{E}) \) is a metric space.

The metric for this space when \( \alpha \in (0, 1] \) is

\[
d_\alpha(z^{(1)}, z^{(2)}) = \sum_{j \in J} E|z_j^{(1)} - z_j^{(2)}|^\alpha,
\]

where \( z^{(1)} = (z_0^{(1)}, z_1^{(1)}, \ldots) \) and \( z^{(2)} = (z_0^{(2)}, z_1^{(2)}, \ldots) \) are arbitrary elements of \( \ell^e_\alpha(\Omega, B, P; \mathbb{E}) \).

Let \( G \) be the basic transition function of \( A \). We say \( G \) satisfies the Lipschitz condition of order \( \alpha \) if there exists a positive constant \( K \) such that \( \max_{\sigma \in \mathbb{E}} |G(\sigma, e) - G(\sigma, e')| \leq K|e - e'|^\alpha \) for all \( e, e' \in \mathbb{E} \). We shall only consider \( \alpha \in (0, 1] \) since other values imply that \( G(\sigma, \cdot) \) is not continuous or that \( G(\sigma, \cdot) \) is a constant for each \( \sigma \in \mathbb{E} \).

Let \( x^{(1)}, x^{(2)}, \ldots \) be a sequence of environment sequences in \( \ell^e_\alpha(\Omega, B, P; \mathbb{E}) \) that converges to \( x \) in \( \ell^e_\alpha(\Omega, B, P; \mathbb{E}) \). That is, \( d_\alpha(x^{(i)}, x) \to 0 \) as \( i \to \infty \). Let us now suppose that \( G \) satisfies the Lipschitz condition of order \( \alpha \). For any \( x \in \mathbb{E}^K \) and any \( \omega \in \Omega \), by lemma 3.2.2. we have

\[
|\prod_{j=1}^k |G(\sigma_j, z_j^{(1)}(\omega)) - G(\sigma_j, z_j(\omega))| \leq K \sum_{j=1}^k |z_j^{(1)}(\omega) - z_j(\omega)|^\alpha.
\]

However, by the Lipschitz condition

\[
\sum_{j=1}^k |G(\sigma_j, z_j^{(1)}(\omega)) - G(\sigma_j, z_j(\omega))| \leq K \sum_{j=1}^k |z_j^{(1)}(\omega) - z_j(\omega)|^\alpha.
\]

Thus

\[
E \prod_{j=1}^k |M(x, z^{(1)}) - M(x, z)| = \int_\Omega \prod_{j=1}^k |M(x, z^{(1)}(\omega)) - M(x, z(\omega))| dP(\omega)
\]

\[
\leq K \sum_{j=1}^k |z_j^{(1)}(\omega) - z_j(\omega)|^\alpha dP(\omega)
\]
\[ K \sum_{j=1}^{k} \int_{\Omega} |z_j^{(1)}(\omega) - z_j(\omega)|^q dP(\omega) \]

\[ \leq K \sum_{j \in J} \int_{\Omega} |z_j^{(1)}(\omega) - z_j(\omega)|^q dP(\omega) \]

\[ = K d_\alpha(z^{(1)}, z). \]

But \( d_\alpha(z^{(1)}, z) \to 0 \) as \( i \to \infty \). Thus, \( E|\mathcal{M}(x, z^{(1)}) - \mathcal{M}(x, z)| \to 0 \) as \( i \to \infty \) for all \( x \in \mathcal{E}^* \). Hence, \( \mathcal{M}(x, z^{(1)}) \to \mathcal{M}(x, z) \) in probability uniformly for all \( x \in \mathcal{E}^* \). Since \( z^{(1)} \to z \) in \( L_\alpha^q(\Omega, \mathcal{B}, P; E) \), we have that \( z_0^{(1)} \to z_0 \) in probability. If \( \pi_0 \) is continuous, then \( \pi_0(z_0^{(1)}) \to \pi_0(z_0) \) in probability. Thus, by the fact that the transition functions are bounded, we have for all \( x \in \mathcal{E}^* \) that \( \pi(x, z^{(1)}) = \pi_0(z_0^{(1)}) \mathcal{M}(x, z^{(1)}) \to \pi_0(z_0) \mathcal{M}(x, z) = \pi(x, z) \) in probability and \( p(x, z^{(1)}) = \pi(x, z^{(1)}) \eta^F \to \pi(x, z) \eta^F = p(x, z) \) in probability as \( i \to \infty \).

Consequently, we have a stability and approximation result under the above conditions.

It is also possible to obtain an ever-ergodic type of stability for automata in random environments.

**Definition 4.3.1.** An ARE \((A, Z)\) is ever-ergodic a.s. iff for any \( \epsilon > 0 \), there exists \( W \subset \Omega \), where \( P(W) = 0 \), and there exists \( N(\epsilon) \) such that \( \mathcal{N}(x, z_k(\omega)) < \epsilon \) \( \forall z_k(\omega) \in \mathcal{E}^* \) whenever \( \omega \notin W \) and \( L(x) \geq N(\epsilon) \).

Note that \( z_k(\omega) \) is the k-tail of \( z(\omega) \). Hence, for every \( \omega \notin W \), \( A \) is ever-ergodic at \( z(\omega) \) and the choice of \( N(\epsilon) \) is independent of \( \omega \).

Suppose there exists \( 0 < \epsilon_1 < 1 \) such that there exists \( W_{\epsilon_1} \subset \Omega \),
where $P(W_\varepsilon) = 0$, and there exists $N(\varepsilon_1)$ whereby $\forall \varepsilon_1 \in (0, \varepsilon_1)$ whenever $\omega \notin W_{\varepsilon_1}$ and $L(x) \geq N(\varepsilon_1)$. For any $\varepsilon$ such that $0 < \varepsilon < \varepsilon_1$, we can find $n_0$ such that $\varepsilon_1 < \varepsilon$. By lemma 3.3.2 we have that $\forall \varepsilon_1 \in (0, \varepsilon_1)$ whenever $\omega \notin W_{\varepsilon_1}$ and $L(x) \geq n_0 N(\varepsilon_1)$. Thus, we can choose $W$ to be independent of $\varepsilon$.

Let $(A, Z)$ be an ARE. If there exists $\delta > 0$ such that each component of $G(\sigma, z_1)$ $\forall \sigma \in \Sigma$ and $\forall \sigma \in J_+$ is greater than $\delta/2$ for almost every $\omega \in \Omega$, then $\exists G(\sigma, z_1) < 1-\delta$ a.s. $\forall \sigma \in \Sigma$ and $\forall \sigma \in J_+$. By application of lemma 3.3.1 we see that $\exists G(\sigma, z_1) < (1-\delta)^n_0$ a.s. $\forall z_1 \in E^*(z)$ and $\forall x \in E^*$ such that $L(x) \geq n_0$. Thus, we have generated a sufficient condition for $(A, Z)$ to be ever-ergodic a.s.

Theorem 4.3.1. Let $(A, Z)$ be an ARE with natural metric $d^N$ on $E$. If $(A, Z)$ is ever-ergodic a.s., $G$ is uniformly continuous and $\pi_0$ is uniformly continuous, then given any $\varepsilon > 0$, there exists $\delta > 0$ such that for any ARE $(A, Z')$, where $Z'$ is defined on $(\Omega, \mathcal{B}, P)$, satisfying $\sup_{\sigma \in J} d^N(z_1(\omega), z'_1(\omega)) < \delta$ for almost every $\omega \in \Omega$, we have $\pi(x, z(\omega)) - \pi(x, z'(\omega)) < \varepsilon$ $\forall x \in E^*$ and for almost every $\omega \in \Omega$.

The proof of theorem 4.3.1 follows in a similar way as corollary 3.3.7 and its antecedents except that the results apply uniformly for almost every $\omega \in \Omega$. Furthermore, if we have a sequence of $z(1), z(2), \ldots$ of random environment sequences such that $z(1) + z$ a.s., then under the continuity conditions of theorem 4.3.1 we obtain that
\( \pi(x,z^{(i)}) \to \pi(x,z) \) a.s. \( \forall x \in \mathcal{E}^* \). And, consequently, \( p(x,z^{(i)}) \to p(x,z) \) a.s. \( \forall x \in \mathcal{E}^* \).

4.4 A Canonical Representation for Automata in Random Environments

In section 4.1 we encountered the dual nature of the randomness involved in automata in random environments. We have that the environment sequence is random and that for each value the environment sequence assumes, the probabilistic state transitions are defined. Hence, the state of the machine after an input tape under a random environment sequence is a function defined on a product probability space. We shall show that for any ARE \((A,Z)\) there exists an ARE \((A_D,Y)\) over the same alphabet whose transition probabilities are equivalent in the mean to those of \((A,Z)\), but the state of \((A_D,Y)\) obtained after an input under a random sequence is only related to the value the random environment sequence obtains. To accomplish this we need that the initial distribution function and the basic matrix transition functions of \(A_D\) take on values which, in addition to being stochastic vectors and stochastic matrices, respectively, must also have components that are either 0 or 1. Thus, the only randomness in the system \((A_D,Y)\) is the randomness of the environment sequence. We shall also construct \((A_D,Y)\) to have a finite environment set.

**Theorem 4.4.1.** Let \((A,Z)\) be an ARE, where \(A = (\Sigma,S,G,\pi_0,F,E)\) is an ADE and \(Z = \{z_j\}_{j \in J}\) is an ESP defined on \((\Omega,\mathcal{B},P)\) taking values in \(E\). There exists an ARE \((A_D,Y)\), where \(A_D = (\Sigma,S,H,\rho_0,F,E')\) is an
ADE and $Y = \{y_j \mid j \in J\}$ is an ESP defined on some $(\Omega', \mathcal{B}', P')$ taking values in $\mathcal{E}'$, such that $(A_D, Y)$ satisfies the following properties:

1. $H: \Sigma \times \Omega' \to D(\mathcal{H}_n)$
2. $\rho_0: \Omega' \to D(\mathcal{V}_n)$
3. $\#(\mathcal{E}') < \infty$
4. $E\pi(x, s) = E\rho(x, y) \forall x \in \Sigma$, where $y = (y_0, y_1, \ldots)$
5. $ET(A, x, \lambda) = ET(A_D, y, \lambda) \forall \lambda \in [0, 1)$.

Here we are using $\pi$ and $\rho$ to denote the state distribution functions of $(A, Z)$ and $(A_D, Y)$, respectively. Recall from definition 1.3.3 that $D(\mathcal{H}_n)$ is the set of all $n \times n$ stochastic matrices with components 0 or 1 and that $D(\mathcal{V}_n)$ is the set of all $n$-dimensional stochastic vectors with components 0 or 1. Note that $n$ is the cardinality of the state set $S$. This property implies that knowledge of the input and environment sequence determines the state of $(A_D, Y)$ obtained.

**Proof:** For any $s \in \Sigma$ and $e \in \mathcal{E}$ we may decompose $G(s, e)$ into a finite convex linear combination of elements of $D(\mathcal{H}_n)$ by the following algorithm:

1. Set $G(s, e) = G_1$ and set $i = 1$.
2. Let $g_{jk}^i$ be the smallest nonzero component of $G_1$, where $j$ and $k$ are its coordinates. If there is a tie, choose any one of the contenders.
3. Set $a_{i0}(e) = g_{jk}^i$.
4. Construct $B_{i0}(e) \in D(\mathcal{H}_n)$ by the rule:
   a. $(B_{i0}(e))_{jk} = 1$. 
b. For r ≠ j, find the largest component of the r-th row of \( G_1 \), say \( g_{rt} \), then set \( (B_{1\sigma}(e))_{rt} = 1 \).

If there is a tie within any row, choose any one of the contenders.

c. Set all of the remaining components of \( B_{1\sigma}(e) \) equal to zero.

5. Set \( G_{i+1} = G_i - \alpha_{1\sigma}(e)B_{1\sigma}(e) \).

6. If \( G_{i+1} \) is a matrix of all 0 components, stop. If not, increase the value of \( i \) by 1 and go to step 2.

This algorithm must terminate in at most \( p = n^2 - n + 1 \) iterations regardless of \( e \) and \( \sigma \) because at each iteration at least one component of \( G_i \) is zeroed and the rows contain nonnegative components summing to \( 1 - \sum_{j=1}^{i} \alpha_{j\sigma}(e) \). Clearly for each \( i \), \( B_{i\sigma}(e) \epsilon D(M_n) \). If for some \( \sigma \) and \( e \) this process terminates before \( i = p \), we define \( \alpha_{i\sigma}(e) = 0 \) for the remaining values of \( i \) and assign the corresponding \( B_{i\sigma}(e) \) any arbitrary value in \( D(M_n) \).

So we now have \( \sum_{i=1}^{p} \alpha_{i\sigma}(e) = 1 \) and \( \sum_{i=1}^{p} \alpha_{i\sigma}(e)B_{i\sigma}(e) = G(\sigma,e) \) for each \( \sigma \epsilon \Sigma \) and each \( e \epsilon E \). Hence for each \( e \epsilon E \), we have generated the following arrays of numbers and matrices:

\[
\begin{pmatrix}
\alpha_{1\sigma_1}(e) & \cdots & \alpha_{p\sigma_1}(e) \\
\vdots & \ddots & \vdots \\
\alpha_{1\sigma_m}(e) & \cdots & \alpha_{p\sigma_m}(e)
\end{pmatrix}
\begin{pmatrix}
B_{1\sigma_1}(e) & \cdots & B_{p\sigma_1}(e) \\
\vdots & \ddots & \vdots \\
B_{1\sigma_m}(e) & \cdots & B_{p\sigma_m}(e)
\end{pmatrix}
\]

where \( m \) is the cardinality of \( E \). For any \( i_1, \ldots, i_m = 1, \ldots, p \), let
\[ a_e(i_1, \ldots, i_m) = \prod_{j=1}^{m} a_{ij} \sigma_j(e). \] We now consider \( A(e) = \{a_e(i_1, \ldots, i_m) \mid i_1, \ldots, i_m = 1, \ldots, p\}^L \) as a list of numbers. A list differs from a set in that redundancy is preserved. That is, the lists \( \{0\}^L \) and \( \{0,0\}^L \) are not equal because 0 occurs once in the first list and twice in the second list. Clearly for any \( e \in \Sigma \) and any \( i_1, \ldots, i_m = 1, \ldots, p, \)
\[ a_e(i_1, \ldots, i_m) \geq 0. \] Now for any \( e \in \Sigma \),
\[
\sum_{i_1} \ldots \sum_{i_m} a_e(i_1, \ldots, i_m) = \sum_{i_1} \ldots \sum_{i_m} \prod_{j=1}^{m} a_{ij} \sigma_j(e)
\]
\[
= \sum_{i_1} \ldots \sum_{i_{m-1}} \prod_{j=1}^{m-1} a_{ij} \sigma_j(e) \sum_{i_m} a_{i_m} \sigma_m(e)
\]
\[
= \sum_{i_1} \ldots \sum_{i_m} \prod_{j=1}^{m-1} a_{ij} \sigma_j(e) = \ldots = 1.
\]

For each \( \sigma_k \in \Sigma \), we may partition \( A(e) \) into \( p \) sublists
\[ A_{1\sigma_k}(e), \ldots, A_{p\sigma_k}(e), \] where
\[ A_{1\sigma_k}(e) = \{a_e(i_1, \ldots, i_{k-1}, i, i_{k+1}, \ldots, i_m) \mid i_1, \ldots, i_{k-1}, i, i_{k+1}, \ldots, i_m = 1, \ldots, p\}^L. \]

We obtain the sum of the elements of the list \( A_{1\sigma_k}(e) \)
\[
\sum_{i_1} \ldots \sum_{i_{k-1}} \sum_{i_{k+1}} \ldots \sum_{i_m} a_e(i_1, \ldots, i_{k-1}, i, i_{k+1}, \ldots, i_m)
\]
\[
= \sum_{i_1} \ldots \sum_{i_{k-1}} \sum_{i_{k+1}} \ldots \sum_{i_m} a_{1\sigma_k}(e) \prod_{j \neq k} a_{ij} \sigma_j(e)
\]
\[
= a_{1\sigma_k}(e) \sum_{i_1} \sum_{i_{k-1}} \sum_{i_{k+1}} \ldots \sum_{i_m} \prod_{j \neq k} a_{ij} \sigma_j(e) = a_{1\sigma_k}(e).
\]
There are $p^m$ elements in the list $A(e)$. We now construct the list $A'(e) = \{a'_i(e)|i=1,\ldots,p^m\}^L$. For $i = 1 + \sum_{j=1}^{m} (i_j-1)p^{j-1}$, set $a'_i(e) = a_{e(i_1,\ldots,i_m)}$. This uniquely defines each element of $A'(e)$ and $A'(e) = A(e)$. For each $\sigma_k \in \Sigma$, we partition $A'(e)$ into $p$ sublists $A'_{1\sigma_k}(e),\ldots,A'_{p\sigma_k}(e)$, where $a'_i(e) \in A'_{j\sigma_k}(e)$ iff 

$$a_{e(i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_m)} \in A'_{j\sigma_k}(e) \quad \text{and} \quad i = 1 + (j-1)p^{k-1} + \sum_{r=1}^{k-1} (i_r-1)p^{r-1}.$$ 

Actually, $A'_{j\sigma_k}(e) = A_{j\sigma_k}(e)$. Thus, we have reordered the set $A(e)$ and maintained the properties of the partitioning.

We now construct a list $B'(e) = \{B'_{i\sigma}(e)|i=1,\sigma \in \Sigma,\ldots,p^m\}^L$ by the rule:

1. Set $i = 1$ and $k = 1$.

2. $a'_i(e)$ must belong to at least one of the sublists $A'_{1\sigma_k}(e),\ldots,A'_{p\sigma_k}(e)$, say $A'_{j\sigma_k}(e)$. Set $B'_{i\sigma}(e) = A'_{j\sigma_k}(e)$.

3. Delete one occurrence of $a'_i(e)$ from the list $A'_{j\sigma_k}(e)$, but let the amended list retain the name $A'_{j\sigma_k}(e)$.

4. Increase the value of $i$ by 1.

5. If $i \leq p^m$, go to step 2. If not, go to step 6.

6. If $k < m$, then set $i=1$, increase the value of $k$ by 1, and go to step 2. If $k \geq m$, stop.

This algorithm must terminate because for each $\sigma \in \Sigma$, $A'_{1\sigma}(e),\ldots,A'_{p\sigma}(e)$ is a partition of the $p^m$ elements of $A'(e)$. Furthermore for each $\sigma \in \Sigma$ each of these sublists will be emptied by the deletion process.

As a result of the construction of $A'(e)$ and $B'(e)$ we obtain for each $e \in E$
1. \( \alpha'_i(e) \geq 0 \ \forall i \in \{1, \ldots, p^m\}, \sum_{i=1}^{p^m} \alpha'_i(e) = 1. \)

2. For each \( \sigma \in \Sigma \), \( A'_i(e) \) can be partitioned into \( p \) sublists
   
   \[ \sum_{i \in A'_i(e) \in \text{sublist}} \alpha'_i(e) = \alpha_j(e). \]

3. \( B'_{i\sigma}(e) \in D(M^n) \ \forall i = 1, \ldots, p^m \) and \( \forall \sigma \in \Sigma \) and \( B'_{i\sigma}(e) = B_{j\sigma}(e) \)
   when \( i \) is an index of an element in \( A'_{j\sigma}(e) \).

Also, for any \( \sigma \in \Sigma \) and \( e \in E \),

\[
\sum_{i=1}^{p^m} \alpha'_i(e)B'_{i\sigma}(e) = \sum_{j=1}^{p} \sum_{i \in A'_{j\sigma}(e)} B_{j\sigma}(e) \alpha'_i(e) \\
= \sum_{j=1}^{p} B_{j\sigma}(e) \sum_{i \in A'_{j\sigma}(e)} \alpha'_i(e) \\
= \sum_{j=1}^{p} B_{j\sigma}(e) \alpha_j(e) = G(\sigma, e).
\]

Hence, for each \( e \in E \) we have the following arrangement

\[
\begin{array}{cccc}
\alpha'_1(e) & B'_{1\sigma_1}(e) & \cdots & B'_{1\sigma_m}(e) \\
\vdots & \vdots & \ddots & \vdots \\
\alpha'_{p^m}(e) & B'_{p^m\sigma_1}(e) & \cdots & B'_{p^m\sigma_m}(e)
\end{array}
\]

Consider now the finite set \( D_{n,m} = D(M^n) \times \cdots \times D(M^n) \), the Cartesian product of \( D(M^n) \) with itself \( m \) times. Since \( D(M^n) \) is a finite set with \( n^n \) distinct elements, then \( D_{n,m} \) is a finite set with \( n^{nm} \) distinct elements. Since \( D_{n,m} \) is a finite set, we can order it in a systematic way. For each \( k = 1, \ldots, n^{nm} \), take the \( k \)-th element of
By our ordering and call it $D_k$. Also, for each $e \in E$ and $k = 1, \ldots, n^{nm}$, define $I_k(e) = \{i \mid (B_i^{\sigma^1}(e), \ldots, B_i^{\sigma^m}(e)) = D_k\}$. Note that $I_k(e) \cap I_{k'}(e) = \emptyset$ for $k \neq k'$ and $\bigcup_{k=1}^{n^{nm}} I_k(e) = \{1, \ldots, p^m\}$.

Thus, we set

$$
\beta_k(e) = \begin{cases} 
\sum_{i \in I_k(e)} a_i^e(e) & \text{if } I_k(e) \neq \emptyset \\
0 & \text{if } I_k(e) = \emptyset 
\end{cases}
$$

Hence, $\beta_k(e) \geq 0$ for $k = 1, \ldots, n^{nm}$ and for all $e \in E$. Also for each $e \in E$,

$$
\sum_{k=1}^{n^{nm}} \beta_k(e) I_k(e) = \sum_{k=1}^{n^{nm}} I_k(e) = \sum_{i=1}^{p^m} a_i^e(e) = 1
$$

and

$$
\sum_{k=1}^{n^{nm}} \beta_k(e) D_k = \sum_{k=1}^{n^{nm}} I_k(e) (B_i^{\sigma^1}(e), \ldots, B_i^{\sigma^m}(e))
$$

$$
= \sum_{i=1}^{p^m} a_i^e(e) (B_i^{\sigma^1}(e), \ldots, B_i^{\sigma^m}(e))
$$

$$
= (G(\sigma^1, e), \ldots, G(\sigma^m, e)).
$$

We are now ready to define $(\mathcal{A}_D, Y)$. We retain $\Sigma, S,$ and $F$ from $(A, Z)$. Let $E' = \{1, \ldots, n^{nm}\}$. Define $H(\sigma_i^j, k)$ to be the $i$-th matrix in $D_k$. For $k = 1, \ldots, n$, let $\rho_0(k)$ be an $n$-dimensional row vector of all zeroes except for a $1$ in the $k$-th component. For $n+1 \leq k \leq n^{nm}$, let $\rho_0(k)$ be the $n$-dimensional vector $(1, 0, \ldots, 0)$. Hence, $H$ is defined for all $\sigma \in \Sigma$ and all $e' \in E'$ and $\rho_0$ is defined for all $e' \in E'$. 
Define \( \gamma_k(e) = \begin{cases} \pi_0^{(k)}(e) & k = 1, \ldots, n \\ 0 & k = n+1, \ldots, n^{nm} \end{cases} \)

where \( \pi_0^{(k)}(e) \) is the \( k \)-th component of \( \pi_0(e) \). By Kolmogorov's Existence Theorem we define the process \( Y = \{y_j | j \in J\} \) by the finite dimensional distributions

\[
\psi(y_{j_1} = k_1, \ldots, y_{j_r} = k_r) = \int_{\Omega} \beta_{k_1} (z_{j_1}) \cdots \beta_{k_r} (z_{j_r}) dP,
\]

where \( j_1, \ldots, j_r \in J \) are all distinct. If some \( j_1 = 0 \), replace \( \beta_{k_1} (z_{j_1}) \) by \( \gamma_{k_1} (z_{j_1}) \). Note that the integral is taken with respect to the probability space defining \( Z \).

Thus, for any \( r \in J \),

\[
P(y_0 = k_0, y_1 = k_1, \ldots, y_r = k_r) = \int_{\Omega} \gamma_{k_0} (z_0) \beta_{k_1} (z_1) \cdots \beta_{k_r} (z_r) dP.
\]

Let \( x \in \Sigma^* \) be an input tape of arbitrary length, say \( x = \sigma_1 \sigma_2 \cdots \sigma_r \). Hence

\[
\mathbb{P}(x, y) = \sum_{k_0} \sum_{k_r} \rho_0(k_0) H(\sigma_1, k_1) \cdots H(\sigma_r, k_r) P(y_0 = k_0, y_1 = k_1, \ldots, y_r = k_r)
\]

where each sum is taken from 1 to \( n^{nm} \). By our construction of \( Y \),

\[
\mathbb{P}(x, y) = \sum_{k_j} \sum_{k_r} \rho_0(k_0) H(\sigma_1, k_1) \cdots H(\sigma_r, k_r) \int_{\Omega} \gamma_{k_0} (z_0) \beta_{k_1} (z_1) \cdots \beta_{k_r} (z_r) dP
\]

\[
= \int_{\Omega} \sum_{k_j} \sum_{k_r} \rho_0(k_0) H(\sigma_1, k_1) \cdots H(\sigma_r, k_r) \gamma_{k_0} (z_0) \beta_{k_1} (z_1) \cdots \beta_{k_r} (z_r) dP
\]

\[
= \int_{\Omega} \left[ \sum_{k_0} \gamma_{k_0} (z_0) \right] \left[ \sum_{k_1} H(\sigma_1, k_1) \beta_{k_1} (z_1) \right] \cdots
\]

\[
\cdots
\]

\[
\cdots
\]

\[
\cdots
\]
\[
\left( \sum_{k_1} H(\sigma_{r,k_1}) \prod_{k_2} (z_{r,k_2}) \right) dP \\
= \int_{\Omega} \pi_0(z_0) G(\sigma_1, z_1) \ldots G(\sigma_r, z_r) dP \quad \text{by construction} \\
= \int_{\Omega} \pi(x,z) dP = E\pi(x,z) .
\]

Moreover, if we denote the probability of acceptance of \(x\) by \((A_0, Y)\) under random environment sequence \(y\) by \(p'(x,y)\), we have

\[
E \pi'(x,y) = E \pi(x,y) \eta^F = (E \pi(x,y)) \eta^F \\
= (E \pi(x,z)) \eta^F = E \pi(x,z) \eta^F = E \pi(x,z) .
\]

Thus, since the expected probability of acceptance remains unchanged,

\[
E(T(A_0, Y, \lambda)) = E(T(A, z, \lambda)) \quad \forall \lambda \in [0,1).
\]

**Example 4.4.1.** Let \(A = (E, S, G, \pi_0, F, E)\), where \(E = \{\sigma_1, \sigma_2\}\), \(S = \{s_1, s_2\}\), and \(E = [0,1]\). Let \(G\) be defined as

\[
G(\sigma_1, e) = \begin{bmatrix} e & 1-e \\ 1-e^2 & e^2 \end{bmatrix} \quad G(\sigma_2, e) = \begin{bmatrix} e^2 & 1-e^2 \\ \sin e & 1-\sin e \end{bmatrix}
\]

Clearly, \(G(\sigma_1, e) = (e-e^2) \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + (1-e) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + e^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \forall e \in E\)

and \(G(\sigma_2, e) = \begin{cases} (e-e^2) \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} + \sin e \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + (e^2-\sin e) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \forall e \in E \text{ such that } e \geq e_0 \\
(1-\sin e) \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} + e^2 \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + (\sin e-e^2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \forall e \in E \text{ such that } e < e_0 ,
\end{cases}\)
where $e_0 \in \mathbb{E}$ such that $e_0 \neq 0$ and $\sin e_0 - e_0^2 = 0$. The approximate value of $e_0$ is .88.

We now define $a_i'(e)$ for $i = 1, \ldots, 6$ to be the functions

<table>
<thead>
<tr>
<th>$a_i'(e)$</th>
<th>$0 \leq e &lt; e_0$</th>
<th>$e_0 \leq e \leq 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1'(e)$</td>
<td>$\frac{e^2}{e^2}$</td>
<td>$\sin e$</td>
</tr>
<tr>
<td>$a_2'(e)$</td>
<td>$\sin e - e^2$</td>
<td>$0$</td>
</tr>
<tr>
<td>$a_3'(e)$</td>
<td>$e - \sin e$</td>
<td>$e - e^2$</td>
</tr>
<tr>
<td>$a_4'(e)$</td>
<td>$1 - e$</td>
<td>$0$</td>
</tr>
<tr>
<td>$a_5'(e)$</td>
<td>$0$</td>
<td>$e^2 - \sin e$</td>
</tr>
<tr>
<td>$a_6'(e)$</td>
<td>$0$</td>
<td>$1 - e$</td>
</tr>
</tbody>
</table>

Hence for all $e \in \mathbb{E}$, $a_i'(e) \geq 0$ and $\sum_{i=1}^{6} a_i'(e) = 1$. We now define $B_{i\sigma}'(e)$ for $i = 1, \ldots, 6$ and $\sigma \in \mathbb{E}$

<table>
<thead>
<tr>
<th>$B_{1\sigma_1}'(e)$</th>
<th>$(1 \ 0 \ 1)$</th>
<th>$B_{1\sigma_2}'(e)$</th>
<th>$(1 \ 0 \ 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{2\sigma_1}'(e)$</td>
<td>$(1 \ 0 \ 0)$</td>
<td>$B_{2\sigma_2}'(e)$</td>
<td>$(1 \ 0 \ 1)$</td>
</tr>
<tr>
<td>$B_{3\sigma_1}'(e)$</td>
<td>$(1 \ 0 \ 0)$</td>
<td>$B_{3\sigma_2}'(e)$</td>
<td>$(0 \ 1 \ 1)$</td>
</tr>
<tr>
<td>$B_{4\sigma_1}'(e)$</td>
<td>$(0 \ 1 \ 1)$</td>
<td>$B_{4\sigma_2}'(e)$</td>
<td>$(0 \ 1 \ 1)$</td>
</tr>
<tr>
<td>$B_{5\sigma_1}'(e)$</td>
<td>$(1 \ 0 \ 0)$</td>
<td>$B_{5\sigma_2}'(e)$</td>
<td>$(0 \ 1 \ 0)$</td>
</tr>
<tr>
<td>$B_{6\sigma_1}'(e)$</td>
<td>$(0 \ 1 \ 0)$</td>
<td>$B_{6\sigma_2}'(e)$</td>
<td>$(0 \ 1 \ 0)$</td>
</tr>
</tbody>
</table>

Clearly, $\sum_{i=1}^{6} a_i'(e)B_{i\sigma}'(e) = G(\sigma, e)$ for all $\sigma \in \mathbb{E}$ and $e \in \mathbb{E}$. 
CHAPTER V
Implications

5.1 Limit Results for Markov Systems.

Automata theory is generally concerned with input tapes of finite length. Because of its application to Markov systems in environments, we shall now allow input tapes that are countable sequences of elements of $\Sigma$. We shall let $\hat{\Sigma}$ denote the set of all countable sequences of elements of $\Sigma$. If $\hat{x} = \sigma_1\sigma_2\ldots\epsilon\hat{\Sigma}$, then let $\hat{x}_k = \sigma_1\sigma_2\ldots\sigma_k$ be the tape of the first $k$ symbols of $\hat{x}$.

**Definition 5.1.1** Let $C_n$ denote the class of all $n \times n$ stochastic matrices for which there is an eigenvector corresponding to the eigenvalue 1 that is a stochastic vector.

Thus, if $A \in C_n$, there exists $v \in V_n$ such that $vA = v$. If $A$ is an ergodic matrix as defined in section 4.2, then $A \in C_n$.

Suppose we have a countable input sequence $\hat{x} \in \hat{\Sigma}$ and an environment sequence $e \in \hat{\Sigma}$. The following theorems give conditions for $H(\hat{x}_k, e)$ to have a limit as $k \to \infty$.

**Theorem 5.1.1.** Let $A = (\Sigma, S, G, \pi_0, F, \hat{E})$ be an ADE. Let $\hat{x} = \sigma_1\sigma_2\ldots\epsilon\hat{\Sigma}$ and $e = (e_0, e_1, \ldots) \in \hat{\Sigma}^\infty$. If

1. $A$ is ever-ergodic at $e$
2. $G(\sigma_i, e_i) \in C_n$ for all $i = 1, 2, \ldots$
3. $\sum_{i \in \mathbb{J}} ||v_{i+1} - v_i|| < \infty$, where $v_i \in V_n$ such that $v_i G(\sigma_i, e_i) = v_i$,

91
then there exists $M(\hat{x}, e)$ such that $\exists M(\hat{x}, e) = 0$ and

$$\lim_{k \to \infty} ||M(\hat{x}_k, e) - M(\hat{x}, e)|| = 0.$$  

**Proof.** The space of $n$-dimensional vectors with norm $||v|| = \sum_{i=1}^{n} |v_i|$ is complete. So by condition 3, there exists an $n$-dimensional vector $v$ such that $||v_i - v|| \to 0$ as $i \to \infty$. Since $U_n$ is a closed subset of the space of $n$-dimensional vectors and since each $v_i \in U_n$, then $v \in U_n$. Let $V_i \in M_n$ be the matrix each of whose rows is $v_i$. Similarly, let $M(\hat{x}, e)$ be the matrix each of whose rows is $v$. Clearly, $\exists M(\hat{x}, e) = 0$. Note that for any $A \in M_n$ with identical rows, $||A|| = ||a||$, where $a$ is any row of $A$.

For notational convenience, we shall denote $\prod_{i=1}^{i=2} G(\sigma_i, e_i)$ by $\prod_{i=1}^{i=2} G_{i_1}^{i_2}$. Thus,

$$||M(\hat{x}_k, e) - M(\hat{x}, e)|| = ||G_{i_1}^{i_2} - \hat{M}(\hat{x}, e)||$$

$$\leq ||G_{i_1}^{i_2} - V_i|| + ||V_i - \hat{M}(\hat{x}, e)||$$

$$\leq ||G_{i_1}^{i_2} - G_{j_1}^{j_2}|| + ||V_j G_{j_1}^{j_2} - V_i|| + ||V_i - \hat{M}(\hat{x}, e)||.$$  

Note that $V_j G_{j_1}^{j_2} = V_j G_{j_1}^{j_2} + \sum_{i=j}^{k-1} (V_i - V_{i+1}) G_{i+1}^{k}$. But $V_k G(\sigma_k, e_k) = V_k$. Therefore,

$$||V_j G_{j_1}^{j_2} - V_i|| = \sum_{i=j}^{k-1} |(V_i - V_{i+1}) G_{i+1}^{k}||$$

$$\leq \sum_{i=j}^{k-1} ||V_i - V_{i+1}|| \cdot G_{i+1}^{k}$$

$$\leq \sum_{i=j}^{k-1} ||V_i - V_{i+1}|| \cdot \exists G_{i+1}^{k}$$

$$\leq \sum_{i=j}^{k-1} ||V_i - V_{i+1}|| \leq \sum_{i=j}^{\infty} ||V_i + V_{i+1} - V_i||.$$
It follows from condition 3 that for \( j \geq \frac{1}{2}(\varepsilon) \) this last expression can be made less than \( \varepsilon/3 \).

Also, \( \|G_{1}^{j-1}v_{j}G_{j}^{k}\| = (G_{1}^{j-1}v_{j}G_{j}^{k}) \leq \|G_{1}^{j-1}v_{j}\| \sum_{i=j}^{k} \mathcal{T}(G(\sigma_{i},e_{i})) \) by lemma 3.3.3. Since \( A \) is ever-ergodic at \( e \), for fixed \( j \) there exists \( N_{1}(\varepsilon) \) such that the above expression can be made less than \( \varepsilon/3 \) for \( k \geq N_{1}(\varepsilon) \).

And finally \( \|v_{k} - M(\hat{x},e)\| = \|v_{k} - v\| \). By condition 3, this expression can be made less than \( \varepsilon/3 \) for \( k \geq N_{2}(\varepsilon) \).

Hence, \( \|M(\hat{x},e) - M(\hat{x},e)\| \) can be made less than \( \varepsilon \) for sufficiently large \( k \).

Note that theorem 5.1.1 remains true if we replace the ever-ergodic condition by the weaker condition: for \( \hat{x} = \sigma_{1}\sigma_{2}...\varepsilon\hat{x} \) and \( e \in E^{\infty} \),
given any \( \varepsilon > 0 \), there exists \( N(\varepsilon) \) such that
\[
\sum_{i=j}^{j+N(\varepsilon)} \mathcal{T}(G(\sigma_{i},e_{i})) < \varepsilon \quad \forall i \in \mathbb{N}.
\]
We may also weaken condition 2 to require that \( G(\sigma_{i},e_{i}) \notin C_{n} \) for only a finite number of times.

**Theorem 5.1.2.** Let \( A = (E,S,F,F,E) \) be an ADE. Let \( \hat{x} = \sigma_{1}\sigma_{2}...\varepsilon\hat{x} \) and \( e = (e_{0},e_{1},... \in E^{\infty} \). If

1. \( G(\sigma_{i},e_{i}) \notin C_{n} \) for all \( i = 1,2,... \)

2. there exists \( v \in V_{n} \) such that \( \|v_{i} - v\| \to 0 \) as \( i \to \infty \), where \( v_{i} \in V_{n} \) and \( v_{i}G(\sigma_{i},e_{i}) = v_{i} \)

3. there exists a constant \( \gamma \) such that \( \sum_{j=1}^{k} \mathcal{T}(G_{j}) \leq \gamma \) for all \( k = 1,2,... \),
then there exists $\mathcal{M}(\hat{x}, e)$ such that $\mathcal{A}(\mathcal{M}(\hat{x}, e)) = 0$ and

$$\lim_{k \to \infty} ||\mathcal{M}(\hat{x}_k, e) - \mathcal{M}(\hat{x}, e)|| = 0.$$ 

**Proof.** Construct $V_1$ and $\mathcal{M}(\hat{x}, e)$ as in theorem 5.1.1. Since

$$\sum_{j=1}^{k} \mathcal{A}(G_j^k) \leq \gamma,$$

we can make $\mathcal{A}(G_{k-1}^k)$ less than $\varepsilon$ for sufficiently large $k$. But by lemma 3.3.2, $\mathcal{A}(G_j^k) \leq \mathcal{A}(G_{j-1}^k) \mathcal{A}(G_{k-1}^k) \leq \mathcal{A}(G_k^k)$. Hence, $\mathcal{A}(G_j^k) \to 0$ as $k \to \infty$ for all $j \in J_+$. 

Now,

$$||M(\hat{x}_k, e) - M(\hat{x}, e)|| = ||G_1^k - M(\hat{x}, e)|| \leq ||G_1^k - V_k|| + ||V_k - M(\hat{x}, e)||.$$ 

But,

$$||G_1^k - V_k|| \leq ||G(\sigma_1, e_1) - V_1|| G_2^k + \sum_{i=2}^{k} ||V_{i-1} - V_i|| G_1^k \leq ||G(\sigma_1, e_1) - V_1|| \mathcal{A}(G_2^k) + \sum_{i=2}^{k} ||V_{i-1} - V_i|| \mathcal{A}(G_1^k).$$

But $||G(\sigma_1, e_1) - V_1|| \leq 2$ and $\mathcal{A}(G_2^k) \to 0$ as $k \to \infty$. Hence, we can make $||G(\sigma_1, e_1) - V_1|| \mathcal{A}(G_2^k)$ less than $\varepsilon/4$ for $k \geq N_1(\varepsilon)$. 

By condition 2, $||V_{i-1} - V_i|| \to 0$ as $i \to \infty$. So given $\varepsilon > 0$, there exists $k(\varepsilon) > 1$ such that $||V_{i-1} - V_i|| \leq \varepsilon/8$ for $i \geq k(\varepsilon)$. 

For all $k \geq k(\varepsilon) + 1$, \(\sum_{i=k+1}^{k} ||V_{i-1} - V_i|| \mathcal{A}(G_1^k) \leq \varepsilon/9.\) Also, we have seen earlier that $\mathcal{A}(G_1^k) \to 0$ as $k \to \infty$ for all $i$. Hence, given $\varepsilon > 0$, there exists $N_2(\varepsilon)$ such that $\sum_{i=2}^{k} \mathcal{A}(G_1^k) \leq \varepsilon/16$. But $||V_{i-1} - V_i|| \leq 2$;\n
so we have $\sum_{i=2}^{k} ||V_{i-1} - V_i|| \mathcal{A}(G_1^k) \leq \varepsilon/8$. Hence, for $k \geq N_2(\varepsilon)$ we have

$$\sum_{i=2}^{k} ||V_{i-1} - V_i|| \mathcal{A}(G_1^k) \leq \varepsilon/4.$$ 

Hence, $||G_1^k - V_k|| \leq \varepsilon/2$ for $k \geq \max(N_1(\varepsilon), N_2(\varepsilon))$. By condition 2, $||V_k - M(\hat{x}, e)|| \leq \varepsilon/2$ for $k \geq N_3(\varepsilon)$. So given any $\varepsilon > 0$, $||M(\hat{x}_k, e) - M(\hat{x}, e)|| \leq \varepsilon$ for $k \geq \max(N_1(\varepsilon), N_2(\varepsilon), N_3(\varepsilon))$. \(\square\)
We may again weaken condition 1 to require that $G(s_1, e_1) \in C_n$ for only a finite number of times.

By these theorems under the appropriate conditions, we have demonstrated a consistent way to define the transition function of $\hat{x} \in \Sigma$ under the environment sequence $e \in E$.

5.2 Applications.

Consideration of automata in environments has allowed us to generalize the finite-state system model to incorporate the various external factors which may influence the behavior of the system. A probabilistic automaton may be constructed to account for the effects of the environments only when the system can be simulated by a probabilistic automaton. Thus, the interesting case for application of the new automaton in environments model is the case that the system cannot be simulated by a probabilistic automaton. In such a case the environment cannot be reduced to a finite set by the induced pseudo-metric or does not have the appropriate semi-group properties. Moreover, we may incorporate a random structure to the occurrence of the environments. This formulation is particularly useful where there is no control over the occurrence of the environments and we can only make probabilistic assumptions about the occurrences of the environments.

A traditional motive for automata models has been the behavior of digital computers. These systems are indeed situated in environments which affect their mechanical and electrical properties and, hence, their output. Lightning storms, power fluctuations, earth
tremors, and late evening personnel are often blamed as the source of errors. We admit that our sophisticated technology has produced digital computer systems which rarely have transitions other than the ones anticipated.

However, there are many less reliable systems which have probabilistic transitions affected to a greater extent by the configuration of the environment. Information transmission systems may be modeled by automata in environments. The inclusion of the environments will allow us to consider the various types of noise which may occur nonhomogeneously during a transmission.

Another application for automata in environments is modeling learning and pattern recognition. The input represents a stimulus which will encourage some desired behavior. The environment may be taken to be some external circumstances. An alternative interesting possibility for the environments would be their use as records of penalties for previous behavior. The transition functions may then incorporate this history to modify the transition probabilities due to reinforcement of desired behavior.

Decision procedures may be modeled by automata with the input symbols representing the data collected and with the environments added to account for the cost of various actions of the system or to account for certain external phenomena which may be controlled or subject to random variation.

See Ash [1], Chandrasekaran and Shen [5], Krylov and Tsetlin [9], Tsetlin [19], Varshavskii and Vorontsova [21], and Yakowitz [22] for references appropriate to this discussion.
5.3 Conclusions and Summary.

We have taken the viewpoint that every system is situated within environments which affect its behavior. Accordingly, we have generalized the Rabin-Paz model of a probabilistic automaton to account for the effects of the environments. The initial state distribution is no longer considered as a constant, but rather as a function of the initial configuration of the environments. Similarly, the subsequent probabilistic transitions become functions not only of the present input, but also the present configuration of the environment.

We first discussed the case of automata with deterministic environment rules \( e \in E^\infty \). That is, \( e = (e_0, e_1, \ldots) \) specifies the configuration of the environment for the initial distribution and the subsequent transitions. Such a system is called an automaton in deterministic environments (ADE). We might have considered the environments as a nuisance, but actually they enable us to construct a model with greater capacity and flexibility.

It is not always necessary to consider the ADE model for systems with environments. Theorem 2.4.1 demonstrates that a system may be simulated by a probabilistic automaton when the environment set is finite. This involves construction of a probabilistic automaton over a larger input alphabet which incorporates the role of the environments. Theorem 2.4.1 led to further simulation results when the environment set can be reduced to a finite set or when the environment has the appropriate semi-group structure preserved by the system. In section 2.3 we find that there are sets of tapes which can be defined by some
ADE but which cannot be defined by any probabilistic automaton. However, it is possible that such ADE's may be simulated by probabilistic automata.

In Chapter III we tackled the stability problem. It is important because of possible perturbations of the environments or some imprecision of the measurements of the environments. In section 3.3 we discussed sufficient conditions for an ADE to be s-stable for general perturbations of the environments or when the perturbations are measured with respect to a natural metric. We needed to extend the notion of quasidefiniteness to derive the ever-ergodic condition for the proofs. Theorem 3.4.1 and corollary 3.4.2 give sufficient conditions for a-stability without requiring s-stability or the ever-ergodic condition. In section 3.5 we explored the application of the stability results when round off procedures were used in measuring the configuration of the environment.

Finally, we considered the case of automata operating in random environments (ARE). In section 4.1 we defined the new model and discussed procedures to estimate the expected properties of the system. In section 4.2 we discovered that when the random environments are independent then the properties in the mean can be found by consideration of a derived ADE. When $\mathcal{E} = \{\sigma\}$ and $z = (z_0, z_1, \ldots)$ is a sequence of IID random variables, this derived ADE is actually a probabilistic automaton; moreover, we found sufficient conditions to compute the asymptotic moments of the components of $M(\sigma^N, z)$.

We observed the dual nature of the randomness of automata in random environments. However, theorem 4.4.1 demonstrates that there is a mean equivalent canonical representation which eliminates the randomness due
to probabilistic transition; for any ARE, we can find another ARE with finite environment set, deterministic assignment of the initial state, and deterministic transitions which has a state distribution equivalent in the mean. Since the ADE is a special case of the ARE, we find that every ADE has a canonical representation as an ARE with the above properties.

There are some interesting and still open problems which merit further investigation. At various points we have assumed that we could determine whether or not a certain cut-point was isolated. Beyond corollary 3.4.3 we have not given any conditions to locate isolated cut-points. In certain regions of the environment set it may be that the transition functions $G(\sigma,\cdot)$ have certain components which remain 0 for the entire region. In such cases there may be weaker conditions needed for stability. There are also many opportunities for application of the new models to various systems as suggested in section 5.2.
BIBLIOGRAPHY


INDEX

acceptance set 5, 10, 19, 72, 73
acceptor 3
actual automaton 15
admissible 22
automaton in a deterministic environment (ADE) 18
automaton in a random environment (ARE) 72
continuous time probabilistic automaton (CTPA) 35
continuity 38, 59
cut-point 9
definite 12, 13
discriminative probabilistic table (DPT) 11
δ-scrambling 56
effective procedure 6
environment 17
environmental stochastic process (ESP) 71
environment set 17
ever-ergodic a.s. 79
ever-ergodic at e 54
ever-\(H_4(\delta)\) condition 56
final states 4
finite automaton (FA) 4
\(H_4\) condition 14
independent, identically distributed (IID) 73
induced pseudo-metric 29
irreducible 29
isolated cut-point 11, 40
k-definite 12, 13
k-suffix of x 12
k-tail of e 52
length of x 6
linear bounded automaton 3
Lipschitz condition 78
list 84
natural metric 37
nondeterministic finite automaton 7
probabilistic automaton (PA) 8
probabilistic cut-point event (PCE) 10
probabilistic table (PT) 11
quasidfinite 14
random environment 71
regular event 5
response interval 64
right invariant equivalence relation 5
round off 69
scrambling 14
semi-group 4, 32
simulate 21
strictly stable (s-stable) 39
strongly equivalent 11
tape 1, 4
tape-acceptance stable (a-stable) 39
total environment 18
transducer 3
Turing machine 1
weakly ergodic (w-ergodic) 51
weakly k-definite 12, 13
### NOTATION

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>4, 8, 18</td>
</tr>
<tr>
<td>( (A, z) )</td>
<td>72</td>
</tr>
<tr>
<td>( B )</td>
<td>71</td>
</tr>
<tr>
<td>( B' )</td>
<td>71</td>
</tr>
<tr>
<td>( C_n )</td>
<td>91</td>
</tr>
<tr>
<td>( d )</td>
<td>29</td>
</tr>
<tr>
<td>( d_\alpha )</td>
<td>78</td>
</tr>
<tr>
<td>( D(M_n) )</td>
<td>9</td>
</tr>
<tr>
<td>( d_N )</td>
<td>37</td>
</tr>
<tr>
<td>( D_{n,m} )</td>
<td>86</td>
</tr>
<tr>
<td>( D(U_n) )</td>
<td>9</td>
</tr>
<tr>
<td>( E )</td>
<td>17</td>
</tr>
<tr>
<td>( e )</td>
<td>18</td>
</tr>
<tr>
<td>( E^\infty )</td>
<td>18</td>
</tr>
<tr>
<td>( e^\infty )</td>
<td>18</td>
</tr>
<tr>
<td>( E(e) )</td>
<td>59</td>
</tr>
<tr>
<td>( E^*(e) )</td>
<td>52</td>
</tr>
<tr>
<td>( e_k )</td>
<td>52</td>
</tr>
<tr>
<td>( E ) ((A, z, \lambda))</td>
<td>73</td>
</tr>
<tr>
<td>( F )</td>
<td>4</td>
</tr>
<tr>
<td>( F_\gamma )</td>
<td>69</td>
</tr>
<tr>
<td>( F_\gamma^\infty )</td>
<td>69</td>
</tr>
<tr>
<td>( f_\gamma(e) )</td>
<td>69</td>
</tr>
<tr>
<td>( f_\gamma(e) )</td>
<td>69</td>
</tr>
<tr>
<td>( G(\sigma, e) )</td>
<td>18</td>
</tr>
<tr>
<td>( G(\sigma, z) )</td>
<td>72</td>
</tr>
<tr>
<td>( I(x, e, \psi) )</td>
<td>73</td>
</tr>
<tr>
<td>( I(x, z, \psi) )</td>
<td>74</td>
</tr>
<tr>
<td>( J )</td>
<td>39</td>
</tr>
<tr>
<td>( J_+ )</td>
<td>39</td>
</tr>
<tr>
<td>( \ell_\alpha(\Omega, \mathcal{B}, \mathcal{P}) )</td>
<td>77</td>
</tr>
<tr>
<td>( L(x) )</td>
<td>6</td>
</tr>
<tr>
<td>( \tilde{M}_n )</td>
<td>8</td>
</tr>
<tr>
<td>( M(x) )</td>
<td>4, 8</td>
</tr>
<tr>
<td>( H(x, e) )</td>
<td>18</td>
</tr>
<tr>
<td>( H(x, z) )</td>
<td>72</td>
</tr>
<tr>
<td>( P )</td>
<td>71</td>
</tr>
<tr>
<td>( p(x) )</td>
<td>9</td>
</tr>
<tr>
<td>( p(x, e) )</td>
<td>19</td>
</tr>
<tr>
<td>( p(x, z) )</td>
<td>72</td>
</tr>
<tr>
<td>( Q(e_0, \ldots, e_k) )</td>
<td>19</td>
</tr>
<tr>
<td>( Q'(e_1, \ldots, e_k) )</td>
<td>19</td>
</tr>
<tr>
<td>( R )</td>
<td>21, 64</td>
</tr>
<tr>
<td>( S )</td>
<td>4</td>
</tr>
<tr>
<td>( s_0 )</td>
<td>4</td>
</tr>
<tr>
<td>( T )</td>
<td>5</td>
</tr>
<tr>
<td>( T(\Lambda) )</td>
<td>5</td>
</tr>
<tr>
<td>( T(\Lambda, \varepsilon, \lambda) )</td>
<td>19</td>
</tr>
<tr>
<td>( T(\Lambda, z, \lambda) )</td>
<td>72</td>
</tr>
<tr>
<td>( \nu_n )</td>
<td>8</td>
</tr>
<tr>
<td>( x )</td>
<td>4</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>91</td>
</tr>
<tr>
<td>( \lambda_k )</td>
<td>91</td>
</tr>
<tr>
<td>( y )</td>
<td>82</td>
</tr>
<tr>
<td>( z )</td>
<td>71</td>
</tr>
<tr>
<td>( \pi(x) )</td>
<td>9</td>
</tr>
<tr>
<td>( \pi(x, e) )</td>
<td>19</td>
</tr>
<tr>
<td>( \pi(x, z) )</td>
<td>72</td>
</tr>
<tr>
<td>( \rho(x, y) )</td>
<td>82</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>4</td>
</tr>
<tr>
<td>( \Sigma^* )</td>
<td>4</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>91</td>
</tr>
</tbody>
</table>